



**Escola Nacional  
de Saúde Pública**

UNIVERSIDADE NOVA DE LISBOA

## **Curso Mestrado em Gestão da Saúde**

### **Avaliação de Desempenho de Modelos de Ajustamento pelo risco**

**Discente: Manuel Jesus Chantre Lima**

**Orientador: Prof. Doutor Carlos Costa**

**Lisboa, Julho de 2019**



**Escola Nacional  
de Saúde Pública**

UNIVERSIDADE NOVA DE LISBOA

## **Avaliação de Desempenho de Modelos de Ajustamento pelo risco**

Dissertação apresentada para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Gestão da Saúde, realizada sob a orientação científica de: Professor Doutor Carlos Costa

## **Agradecimentos**

Começo por agradecer ao Professor Doutor Carlos Costa pelo seu contributo desde o primeiro momento, no esclarecimento de dúvidas que tornaram possível o planeamento e desenvolvimento deste trabalho. Quero também deixar um agradecimento à minha família, aos meus irmãos e em especial a minha sobrinha Célia pelo apoio e motivação que se revelaram elementos decisivos para o fim desta longa e difícil caminhada. Deixo também um agradecimento especial ao Pedro Piloto que na qualidade de técnico de estatística do Hospital de S. José deu um contributo essencial no processamento da base de dados.

Por último deixar um agradecimento a todos que no âmbito profissional contribuíram para que pudesse conciliar a minha profissão como médico e a realização deste trabalho.

# Índice

<b>1-Introdução.....</b>	<b>1</b>
<b>1.1-Produção Hospitalar .....</b>	<b>4</b>
1.1.1-Identificação e definição da produção hospitalar .....	5
1.1.2 Ajustamento pelo risco .....	7
1.1.3 Severidade, Gravidade e Complexidade.....	8
1.1.4-Dimensões do risco .....	12
1.1.5. Suporte de recolha de dados.....	17
1.1.6- Análise da qualidade dos dados e intervenções corretivas .....	20
1.1.7 - Técnicas de modelação.....	21
1.1.8 - Validade e confiabilidade dos modelos de ajustamento pelo risco .....	22
<b>1.2 - Sistemas de classificação de doentes .....</b>	<b>24</b>
1.2.1 - Sistemas genéricos de classificação de doentes .....	24
1.2.2 - Sistemas de classificação de doentes clínicos .....	32
<b>1.3 - Avaliação do desempenho dos modelos de ajustamento pelo risco .....</b>	<b>36</b>
<b>1.4 - Mortalidade .....</b>	<b>43</b>
1.4.1 - Mortalidade nos DRG e Disease Staging .....	44
1.4.2 - Mortalidade no APACHE II e SAPS II .....	46
<b>2.Objetivos .....</b>	<b>47</b>
<b>3 - Metodologia.....</b>	<b>48</b>
3.1 - Introdução a metodologia do estudo.....	48
3.2 - Fonte de dados .....	49
3.3 - População em Estudo e Critérios de Exclusão .....	49
3.4 – Variáveis em Estudo .....	50
3.4.1- Definições concetuais .....	51
3.4.2 - Definições operacionais .....	55
3.5 - Metodologia e Operacionalização .....	56
<b>4-Resultados.....</b>	<b>60</b>
4.1 Caraterização dos internamentos .....	60

<b>5-Discussão .....</b>	<b>81</b>
<b>5.1 - Discussão metodológica .....</b>	<b>Erro! Marcador não definido.</b>
<b>5.1.1 - Fonte de dados .....</b>	<b>Erro! Marcador não definido.</b>
<b>5.1.2 – Abordagem .....</b>	<b>83</b>
<b>5.1.3 - Limitações operacionais .....</b>	<b>84</b>
<b>5.2 - Discussão de resultados.....</b>	<b>Erro! Marcador não definido.</b>
<b>6. CONCLUSÕES .....</b>	<b>89</b>
<b>7. REFERÊNCIAS BIBLIOGRÁFICAS .....</b>	<b>93</b>
<b>8-ANEXOS .....</b>	<b>124</b>

## Lista de tabelas

<b>Tabela 1</b>	Informação sobre Sistemas
<b>Tabela 2</b>	Informação sobre Modelos Clínicos
<b>Tabela 3</b>	Critérios considerados
<b>Tabela 4</b>	Definições operacionais das variáveis em estudo
<b>Tabela 5 e 6</b>	Estatísticas Descritivas por nível de gravidade para variáveis com interesse clínico. APACHE II
<b>Tabela 7</b>	Estatísticas distribuídas por nível de gravidade para a Mort. Esperada ME e a Mort. Observada MO. APACHE II
<b>Tabela 8 e 9</b>	Estatísticas Descritivas por nível de gravidade para variáveis com interesse clínico. SAPS II
<b>Tabela 10</b>	Estatísticas distribuídas por nível de gravidade para a Mort. Esperada ME e a Mort. Observada MO. SAPS II
<b>Tabela 11 e 12</b>	Estatísticas Descritivas por nível de severidade para variáveis com interesse clínico. APR-DRG
<b>Tabela 13</b>	Estatísticas distribuídas por nível de severidade para a Mort. Esperada ME e a Mort. Observada MO. APR-DRG
<b>Tabela 14</b>	Estatísticas distribuídas por nível de mortalidade para a Mort. Esperada ME e a Mort. Observada MO. APR-DRG
<b>Tabela 15</b>	Estatísticas distribuídas por GDH's para a Mort. Esperada ME e a Mort. Observada MO. APR-DRG
<b>Tabela 16</b>	Distribuição dos casos por níveis de severidade e mortalidade. APR-DRG
<b>Tabela 17</b>	Porcentagem de casos por níveis de severidade e mortalidade. APR-DRG
<b>Tabela 18 e 19</b>	Estatísticas Descritivas por nível de gravidade para variáveis com interesse clínico. D.Staging
<b>Tabela 20</b>	Estatísticas distribuídas por nível de gravidade para a Mort. Esperada ME e a Mort. Observada MO. D.Staging
<b>Tabela 21</b>	Estatísticas distribuídas por BSY para a Mort. Esperada ME e a Mort. Observada MO. APACHE II
<b>Tabela 22</b>	Comparação entre os níveis de severidade e mortalidade do APR-DRG com os níveis de gravidade do D.Staging
<b>Tabela 23</b>	Tempo de internamento vs. Percentis dos 4 Sistemas
<b>Tabela 24</b>	Comparação entre Odds Ratios, % de concordância e acertos na estimativa de óbitos para os 4 Modelos.
<b>Tabela 25</b>	Resultados obtidos para os 4 Modelos
<b>Tabela 26</b>	Episódios identificados como classificações incorretas nos Modelos por Sistema

## Lista de Gráficos

<b>Gráfico 1</b>	Comparação através da razão entre ME e MO por níveis de gravidade para o APACHE II e SAPS II
<b>Gráfico 2</b>	Comparação através dos Odds Ratios por níveis de gravidade para o APACHE II e SAPS II
<b>Gráfico 3</b>	Comparação através da Taxa de Mortalidade por níveis de gravidade para o APACHE II e SAPS II
<b>Gráfico 4</b>	Distribuição da variável GDH APR31 para o APR-DRG
<b>Gráfico 5</b>	Comparação entre níveis de severidade e mortalidade através da razão entre ME e MO. APR-DRG
<b>Gráfico 6</b>	Comparação da Taxa de Mortalidade entre níveis de severidade e mortalidade. APR-DRG
<b>Gráfico 7</b>	Médias do tempo de internamento por Sistema para os 4 percentis
<b>Gráfico 8</b>	Porcentagem de óbitos verificados por percentil em cada um dos 4 Sistemas
<b>Gráfico 9</b>	Distribuição da percentagem de óbitos verificados por percentil em cada um dos 4 Sistemas
<b>Gráfico 10</b>	Comparação entre Odds Ratios, % de concordância e acertos na estimativa de óbitos para os 4 Modelos.
<b>Gráfico 11</b>	Comparação entre a Probabilidade de Morte esperada para os resultados dos 4 Modelos.
<b>Gráfico 12</b>	Distribuição da variável Género
<b>Gráfico 13</b>	Distribuição da variável Tipo de Admissão
<b>Gráfico 14</b>	Distribuição da variável Origem do Doente
<b>Gráfico 15</b>	Distribuição da variável Tipo de Doente
<b>Gráfico 16</b>	Distribuição da variável Motivo de Admissão
<b>Gráfico 17</b>	Percentis vs. Outcome. APACHE II
<b>Gráfico 18</b>	Percentis vs. Outcome. SAPS II
<b>Gráfico 19</b>	Percentis vs. Outcome. APR-DRG
<b>Gráfico 20</b>	Percentis vs. Outcome. D.Staging

## Siglas e Acrónimos

<b>AIM</b>	<i>Acuity Index Method</i>
<b>APACHE</b>	<i>Acute Physiology and Chronic Health Evaluation</i>
<b>AP-DRGs</b>	<i>All Patient Diagnosis Related Groups</i>
<b>APR-DRGs</b>	<i>All Patient Refined Diagnosis Related Groups</i>
<b>CID-9-MC</b> Modificações Clínicas	Classificação Internacional das Doenças – 9ª – Revisão –
<b>CID-10-MC</b> Modificações Clínicas	Classificação Internacional das Doenças – 10ª – Revisão
<b>CSI</b>	<i>Computerized Severity Index</i>
<b>CSP</b>	<i>Complications Screening Program</i>
<b>DGS</b>	Direção Geral de Saúde
<b>DRG</b>	<i>Diagnosis Related Groups</i>
<b>DS</b>	<i>Disease Staging</i>
<b>DXCATS1</b>	Doenças Principal ( <i>Disease Staging</i> )
<b>GCD</b>	Grandes Categorias Diagnósticas (DRGs)
<b>HCFA</b>	<i>Health Care Financing Administration</i>
<b>HCIA</b>	<i>Health Care Information Agency</i>
<b>ICM</b>	Índice de Casemix
<b>IR-DRGs</b>	International Refined Diagnosis Related Groups
<b>MedisGroups</b>	<i>Medical Illness Severity Grouping System</i>
<b>ME</b>	Mortalidade Esperada
<b>MO</b>	Mortalidade Observada
<b>MPM</b>	<i>Mortality Probability Models</i>
<b>NACHRI</b> <i>Institutions</i>	<i>National Association of Children`s Hospitals and Related</i>
<b>Nemtot</b>	Mortalidade prevista recalibrada aos dados portugueses
<b>NYDH</b>	<i>New York State Department of Health</i>
<b>NHS</b>	<i>National Health Service</i>
<b>PDXCAT</b>	Doença Principal ( <i>Disease Staging</i> )
<b>PMCs</b>	<i>Patient Management Categories</i>
<b>PRISM</b>	<i>Pediatric Risk of Mortal</i>
<b>R-DRGs</b>	<i>Refined Diagnosis Related Groups</i>
<b>ROC</b>	<i>Receiver Operating Characteristics Curves</i>
<b>SAPS</b>	<i>Simplified Acute Physiology Score</i>
<b>S-DRGs</b>	<i>Severity Diagnosis Related Groups</i>



<b>SIDA</b>	Síndrome de Imunodeficiência Adquirida
<b>SNS</b>	Serviço Nacional de Saúde
<b>ST1</b>	Estádio 1
<b>3M HIS</b>	<i>3M Health Information Systems</i>
<b>UCI</b>	Unidades de Cuidados Intensivos
<b>VIH</b>	Vírus da Imunodeficiência Humana
<b>3M HIS</b>	3M Health Information Systems

## **Resumo**

**Contexto:** A mortalidade adequadamente ajustada pelo risco é um indicador de efetividade dos cuidados prestados e indesejado para o doente, prestadores de cuidados de saúde, gestores e a sociedade em geral. Também é utilizada para comparar desempenho entre diversos prestadores de cuidados de saúde. Esta última aptidão é obtida com operacionalização dos sistemas de classificação de doentes que atribuem aos doentes um determinado risco de morte com base nas conceções de produto hospitalar e dimensões do risco. No presente estudo foram utilizados quatro sistemas de classificação de doentes, dois genéricos (APR-DRG e Disease Staging) e dois clínicos (APACHE II e SAPS II).

Do estudo conclui-se que os sistemas com significância clínica, como Disease Staging, APACHE II e SAPS II têm maior sensibilidade na previsão da mortalidade e estão relacionados com a procura de cuidados de saúde

**Palavras-chave:** severidade; doente; nível; mortalidade

## Abstract

The study had as general objective to evaluate the adjustment of the four risk adjustment methods (APR-DRG, Disease Staging APACHE II and SAPS II) which measure the severity of the disease and the probability of hospital mortality.

In the specific objectives we aimed to analyze if the models had similar expected mortality; analyze differences and similarities; to analyze the relationship between disease severity and mortality in the four systems.

The population under study: discharge summaries, hospitalization episodes of the Centro Hospitalar Lisboa Central-EPE and internal data from UUM and UCIP-HSJ.

The evaluation of the adjustment of the 4 models as well as their variants according to the variables introduced from the literature review, revealed a good calibration and with significant results for all systems except SAPS II.

Regarding the relationship between the probability of death estimated for the 4 systems given by the percentiles and verified effective deaths, it should be noted that in all systems there is a dependence relation between the variables ( $p < 0.01$ ). In all systems, there is a relationship between the two variables, with the highest percentile being the highest percentage of deaths. The systems with the highest percentages of deaths in the last percentile are SAPS II and Disease Staging, curiously in equal percentage. Disease Staging is the only model in which, in the third percentile, there is a higher percentage of deaths compared to patients classified as discharge alive from hospital. The two generic systems have the lowest percentages of deaths in the first and second percentiles. More results can be checked in detail from the charts.

From the study, it was concluded that systems with clinical significance, such as Disease Staging, APACHE II and SAPS II, are more sensitive in the prediction of mortality and are related to the demand for health care.

Keywords: severity; patient; level; mortality

## 1 Introdução

Os hospitais possuem enorme complexidade e especificidade quer na estrutura quer no seu modelo de administração. Como consequência da complexidade e da especificidade existe uma grande dificuldade na adoção de métodos para avaliar a produção (Bentes e Urbano, 1990; Costa e Lopes, 2004).

Produção hospitalar é indicada pelos casos tratados e baseia-se em dois aspetos: a definição de sistemas de classificação de doentes como metodologia para identificar produtos e a criação de índices de casemix para comparação desses mesmos produtos (Costa e Lopes 2004). Nos sistemas de classificação de doentes estão subjacentes critérios que permitam agregar os doentes tratados de cada hospital (Hornbrook, 1982; Costa, 2004).

Antes de fazer comparações de resultados entre diferentes doentes é importante fazer o ajustamento pelo risco porque ajuda a contextualizar os resultados observados e fornece alguma garantia de que as diferenças observadas estão relacionadas com o estado de saúde prévio dos doentes (Lane-Fall, 2013).

Na definição e elaboração de estratégias adequadas de ajustamento pelo risco, devem ser especificados (Iezzoni LI, 2012): (1) risco de qual resultado; (2) em que período; (3) para que população; (4) para que finalidade. A finalidade determina o quanto bem o ajustador de risco deve executar para ter sucesso (isto é, produzir comparações válidas).

O presente estudo tem como objetivo geral avaliar quatro sistemas de classificação de doentes, na medição da gravidade da doença e da probabilidade de mortalidade hospitalar, em dados do Centro Hospitalar Lisboa Central, EPE.

Para a sua operacionalização vão ser utilizados dois modelos genéricos — APR-DRGs e *Disease Staging* — e dois modelos clínicos — *Acute Physiology and Chronic Health Evaluation* (APACHE) II e *Simplified Acute Physiology Score* (SAPS) II.

Nos objetivos específicos pretende-se comparar a mortalidade observada com os níveis de mortalidade preconizados pelos APR-DRG e com a escala preditiva de mortalidade do *Disease Staging*; comparar a mortalidade observada com níveis de mortalidade preconizados pelo APACHE II e SAPS II; analisar as diferenças entre os quatro sistemas de classificação de doentes na atribuição de probabilidade de morte por níveis de mortalidade e níveis de

gravidade; identificar e analisar possíveis fatores responsáveis pelas diferenças na atribuição da probabilidade de morte; analisar as diferenças entre os dois sistemas de classificação de doentes na atribuição do risco de morte por Grupo de diagnóstico homogêneo (GDH); Doença principal.

A finalidade do estudo é sistematizar e disponibilizar informação que permita um conhecimento mais profundo do desempenho dos quatro sistemas que medem a severidade da doença do doente e o risco de morte

### Estrutura

O presente trabalho encontra-se estruturado em nove capítulos: (1) introdução, onde se apresenta o âmbito e a importância do tema em estudo e a estrutura do trabalho; (2) enquadramento teórico, onde consta o estado da arte referente à problemática em análise; (3) objetivos; (4) metodologia; (5) resultados; (6) discussão, onde se tecem as principais considerações e críticas à metodologia utilizada e aos resultados obtidos; (7) conclusões, onde se encontra a súmula das principais conclusões e relações; (8) bibliografia.

#### **1.1. Produção Hospitalar**

Os hospitais possuem enorme complexidade e especificidade quer na estrutura quer no seu modelo de administração. Como consequência da complexidade e da especificidade existe uma grande dificuldade na adoção de métodos para avaliar a produção (Bentes e Urbano, 1990; Costa e Lopes, 2004), que também é explicada pelas razões associadas tanto à oferta como a procura. (Hornbrook 1982).

Costa e Lopes (2004) acrescentam “o carácter multiproduto da sua atividade, decorrente de uma enorme diversidade nos diagnósticos/doenças que podem ser tratados no hospital, a qual, por sua vez, pode ainda ser potenciada pelo diferente grau da evolução da doença presente no momento de contacto com o hospital”.

Os fundamentos conceituais do produto hospitalar devem ser especificados em concordância com as três perguntas: as preferências que contam, na especificação dos atributos relevantes da produção hospitalar; na perspetiva de teoria económica que pode ser considerado de duas maneiras: 1º, resulta de um processo particular de produção (isto é, tratamento); 2º, a produção é o objeto de troca entre a empresa (hospital) e o consumidor (doente), isto é, produz satisfação ao doente. O hospital é uma estrutura complexa que fornece

uma série interdependente de serviços como, hotelaria, de diagnósticos e terapêuticos. O que é relevante para a empresa? (Berki, 1972).

### **1.1.1. Identificação e definição da produção hospitalar**

A caracterização e a medição da produção hospitalar são decisivas para o conhecimento do que se produz (quer em termos quantitativos quer qualitativos), de como e quem produz, para o apuramento dos custos de produção e para a identificação de elementos que permitam o conhecimento do funcionamento e do desempenho dos hospitais. Portanto, é importante para a investigação e gestão (Bentes, 1996; Costa, 2004).

O produto hospitalar pode ser definido como o conjunto específico de outputs que cada doente recebe em função das suas necessidades e como parte do processo de tratamento definido pelo médico” (Bentes e Urbano, 1988).

Na escolha da medida a utilizar o critério fundamental é o objetivo da medição. Por exemplo, o desenvolvimento de um sistema de pagamento para hospitais exige um método de *casemix* que define grupos de casos que são semelhantes em relação a consumo de recursos hospitalar. Para efeitos de investigação sobre o número, dimensão e especialização dos hospitais, são necessárias medidas de diferenciação de casos para distinguir o grau de especialização e o nível de complexidade dos tipos de doenças e patologias tratadas pelos diferentes hospitais (Hornbrook, 1982; Bentes, 1996; Costa, 2004).

A produção hospitalar acolhe três fundamentais: (1) volume que refere ao total de doentes tratados; (2) *casemix* conjunto de diferentes doenças e problemas de saúde tratados (Fetter et al., 1980; Gonnella, Hornbrook e Louis, 1984); e (3) qualidade é a contribuição dos hospitais para um resultado bem-sucedido ou resolução da doença dos doentes ou problemas de saúde. (Hornbrook, 1985).

É lícito afirmar que a medição do *casemix* é uma questão de agregação e diferenciação que agrupa doenças em "clusters" que são suficientemente semelhantes em relação ao propósito para o qual a medida é necessária, como custo (na conceção de um sistema de pagamento) e divisão da doença de acordo com diferenças significativas na sua natureza que se relacionam com o objetivo principal. O desenvolvimento de medidas de *casemix* implica a derivação de meios alternativos de agrupamento de casos em classes significativas, que

devem então ser avaliadas quanto à consistência com o objetivo de mensuração, seja para pesquisa, gestão ou pagamento (Hornbrook, 1985).

O casemix tem duas dimensões independentes que explicam o custo hospitalar por caso - concentração e gravidade. Concentração é a medida de casemix desenvolvida por Even e Walker (1972) que dá informação sobre o grau de casos tipo, que um hospital trata e que habitualmente tende a concentrar-se em alguns hospitais. O tratamento destes casos exige instalações e equipas especiais. A gravidade reflete a probabilidade de morte, calculada a partir das taxas de mortalidade hospitalar.

A utilização de sistemas de classificação de doentes pressupõe a definição de critérios que permitam agregar os doentes tratados em cada hospital, enquanto que com os índices escalares (Índices de *Casemix*) se procura sintetizar num único indicador a atividade do hospital (Hornbrook, 1982; Costa, 2004) o que é uma vantagem. No entanto, esta abordagem padece de alguns problemas como (Hornbrook, 1982):

- 1- Perda de informação e de sensibilidade relacionada com o nível de agregação;
- 2- Possibilidade de existência de diferentes índices escalares sobre a atividade do hospital, em função da utilização de diferentes dimensões para se atribuírem pesos específicos (por exemplo, complexidade ou gravidade dos casos tratados);
- 3- Possibilidade de não serem adequados para o financiamento dos hospitais, especialmente se refletirem as semelhanças e as diferenças entre os hospitais;
- 4- Desajustado para a avaliação dos cuidados prestados, devido ao nível de agregação.

Segundo citações de Costa e Lopes (2004) referindo a vários autores, existem diversas dimensões da produção hospitalar que podem servir de base para se estabelecerem os pesos específicos:

- 1- A demora média por produtos dos hospitais (Schweiker, 1982);
- 2- Os custos por produtos dos hospitais (Petengill e Vertrees, 1980 e 1981);
- 3- A concentração/especialização dos hospitais (Evans e Walker, 1972; Horn e Schumaker, 1979; Barer, 1982; McGuire e Williams, 1986; Farley e Hogan, 1990);
- 4- A gravidade dos casos tratados (Lichtig, 1986; McGuire, 1991).

### 1.1.2. Ajustamento pelo risco

O ajustamento pelo risco visa explicar as diferenças dos riscos intrínsecos à saúde que os doentes ou as populações são portadores quando contactam os cuidados de saúde. Portanto, é essencial na avaliação de resultados no mundo real que são dependentes dos atributos de risco intrínsecos do doente, efetividade do tratamento, qualidade dos cuidados e probabilidade aleatória. (Iezzoni 2012).

Conceptualmente, o risco corresponde a um acontecimento futuro, incerto, para o qual se conhece a probabilidade de vir a acontecer (Pereira, 2004). É um valor estimado da probabilidade de ocorrência de um dano e a gravidade de tal dano é eminentemente coletivo. Define o todo, mas cada indivíduo distingue-se pela probabilidade de risco que lhe cabe partilhar (Mendes, 2002).

O ajustamento pelo risco consiste num processo estatístico utilizado para identificar e ajustar a variação de resultados dos doentes, originada pela diferença das suas características, entre organizações prestadoras de cuidados de saúde. Tem aplicações na gestão organizacional, na afetação de recursos ou mesmo na investigação (Hornbrook e Goodman, 1996).

Na definição e elaboração de estratégias adequadas de ajustamento pelo risco, devem ser especificados (Iezzoni LI, 2012): (1) risco de qual resultado; (2) período; (3) população; (4) finalidade. Seguem-se outras questões, como, quais os fatores de risco? qual fonte de dados? quais métodos analíticos a serem utilizados?

Existem métodos de ajustamento ao risco que avaliam a diferença de risco (isto é, usam fatores de risco diferentes ou os mesmos fatores de risco de forma diferente), dependendo do resultado alvo (por exemplo, resultado clínico, consumo de recursos ou resultado centrado no doente). Alguns métodos vêm em várias versões, cada versão está calibrada para prever um resultado específico. (Iezzoni, 2012).

Os riscos devem ser enquadrados numa janela temporal, porque clarificam o resultado de interesse e alvitram quais os fatores de risco mais importantes. Os ajustadores ao risco



utilizados na identificação de défices de qualidade visam apreender fatores de risco que antecedem os cuidados (Iezzoni 1994).

A população de interesse ajuda a determinar a série de fatores de risco necessários para avaliar o resultado especificado dentro do prazo pertinente. Por exemplo, na avaliação das taxas de mortalidade das UCIs, os parâmetros fisiológicos relevantes variam entre recém-nascidos, crianças e adultos, embora os achados agudos sejam particularmente relevantes. (Iezzoni, 1994).

A finalidade determina quão bem o ajustador de risco deve executar para ter sucesso (isto é, produzir comparações válidas). Por exemplo, os métodos projetados para prever custos durante um ano, dependem de dados administrativos, que muitas vezes não existem ou são insuficientes e contêm informações clínicas limitadas. Outra finalidade para o ajustamento ao risco é motivar a melhoria da qualidade. Sem esses ajustamentos, os clínicos ou as instituições com resultados maus poderiam argumentar que são tratados injustamente: "os nossos doentes são mais doentes". Algumas vezes para que um ajustador de risco seja percebido, como tal, são necessárias a colheita de dados adicionais e a revisão aprofundada da lógica clínica subjacente ao modelo pelos clínicos participantes (Iezzoni 2012).

Nenhum ajustador de risco é perfeito e nem é necessário, nem possível o ajustamento para todas as características do doente. Portanto, os esforços são dirigidos para identificar fatores de risco que são suficientemente válidos para o propósito explicitado (Iezzoni, 2012).

### **1.1.3. Severidade, Gravidade e Complexidade**

A "gravidade" da doença ou a "severidade" do estado de saúde do doente constituem conceitos "*proxy*", utilizados, muitas vezes, inadequadamente.

Thomas e Longo citados por Costa (2005) identificam como definições de severidade "probabilidade de morte", "diminuição das funções", "carga da doença", "integridade fisiológica ou dificuldade de repor o nível de saúde dos doentes".

McMahon e Billi (1988) citados por Costa (1991) distinguem três perspectivas distintas para a severidade: a fisiológica, a psicológica e a económica. A primeira, procura classificar os doentes em função da probabilidade de morte. A segunda, integra elementos inerentes às respostas emocionais dos doentes face à doença. A terceira, considera o consumo de recursos, ou seja, aquilo que para muitos será «ajustar pela severidade» é sinónimo de «ajustar pela intensidade de recursos».

Os métodos utilizados para medir a gravidade incluem a priori julgamento médico para definir níveis de gravidade dentro de uma doença e a avaliação subjetiva ou a medição fisiológica da condição do doente individual num ponto específico (geralmente admissão ou alta). Os doentes mais graves são aqueles que têm maior probabilidade de morte ou deficiência e / ou cuja gestão (diagnóstico e tratamento) é mais difícil ou complicado comparativamente aos outros. (Yong WW, 1984).

O conceito de complexidade adquire diferentes significados consoante os “atores”. Para os clínicos, complexidade, significa que os doentes tratados têm maior gravidade da doença, apresentam maior dificuldade de tratamento, têm piores prognósticos e requerem maior intervenção. Por outro lado, administradores e reguladores, costumam usar o conceito de complexidade para indicar que os doentes tratados consomem mais recursos, o que resulta num custo maior de prestação de cuidados. (Costa e Lopes, 2004; Hornbrook, 1982).

Existe uma enorme variedade de sistemas de classificação que permitem determinar a severidade da doença dos doentes de hospitais de agudos internados, como são os casos do *Acuity Index Method* (AIM) os *All Patient Refined DRGs* (APR-DRGs), o *Severity of Illness Index* (SII), o *Computerized Severity of Illness* (CSI), o *Disease Staging* (DS) os *International Refined DRGs* (IRDRGs) e os *Medis Groups* (Thomas, Ashcraft e Zimmerman, 1986 e Iezzoni, 1997b). Para as unidades de cuidados intensivos existem igualmente alguns sistemas de classificação validados, como o APACHE II, III e IV (Wong 1991; Knaus 1985; Zimmerman 1998 e 2006), o Mortality Probability Model (MPM II) (Lemeshow et al., 1993), o SAPS II (Le Gall, Lemeshow e Saulnier, 1993) e SAPS III (Moreno, 2005).

**TABELA 1 INFORMAÇÃO SOBRE OS SISTEMAS**

SISTEMA	DEFINIÇÃO DE GRAVIDADE / ABORDAGEM NA CLASSIFICAÇÃO	COMENTÁRIOS
Medições baseadas em Códigos		
<i>Acuity Index Method (AIM)</i>  <b>1. Resultado</b>  <b>2. Tempo de internamento</b>	Tempo de internamento nos GDH  Resultado integral de 1 a 5  Tempo de internamento (dias)	Complicações intra-hospitalares podem contribuir para os resultados; prováveis limitações em relação à avaliação de qualidade.
<i>All Patient Refined Diagnosis-Related Groups (APR-DRGs)</i>	Custos hospitalares nos GDH adjacentes	Complicações intra-hospitalares podem contribuir para os resultados; prováveis limitações em relação à avaliação de qualidade.
<i>Índice de Comorbilidades</i>	Risco de morte em 1 ano após hospitalização médica (resultado integral aditivo)	Baseado apenas em comorbilidades crónicas e não agudas; pode não predizer o desfecho intra-hospitalar.
<i>Disease Staging</i>  <b>1. Probabilidade de morte</b>  <b>2. Tempo de internamento</b>  <b>3. Escala de necessidade de recursos</b>	Probabilidade de morte intra-hospitalar (0 a 1)  Tempo de internamento relativamente a todos os doentes (média = 100)  Uso de recursos em relação a todos os doentes (média = 100)	Complicações intra-hospitalares podem contribuir para os scores; prováveis limitações em relação à avaliação de qualidade.
<i>Patient Mngmt. Categories (PMCs)</i>  <b>1. Tempo de internamento</b>  <b>2. Scores de Intensidade dos Recursos</b>	Tempo de internamento (dias)  Uso dos recursos hospitalares (média = 1.0)  Mortalidade/morbilidade intra-hospitalar (níveis 1 a 7)	Complicações intra-hospitalares podem contribuir para os scores; prováveis limitações em relação à avaliação de qualidade.

<b>3. Nível de gravidade</b>		
<i>Refined Diagnosis-Related Groups (R-DRGs)</i>	Tempo de internamento, charges hospitalares totais nos GDH (grupos B, C e D para doentes médicos; A, B, C, D para doentes cirúrgicos)	Mortes precoces (< 48 horas) são classificadas separadamente. Complicações intra-hospitalares podem contribuir para os scores; prováveis limitações em relação à avaliação de qualidade.
Medições baseadas em registos médicos		
<i>Acute Physiology and Chronic Health Evaluation II (APACHE II)</i>  <b>1. Acute physiology score</b>  <b>2. APACHE</b>	Mortalidade intra-hospitalar nos doentes de cuidados intensivos (resultado integral 0-60)  Inclui o <i>acute physiology score</i> mais os resultados compostos da idade e doença major (resultado integral 0-71).	Desenvolvido e validado para doentes de cuidados intensivos. Scores baseados na condição clínica nas primeiras 24h após a admissão.
<i>Computerized Severity Index (CSI)</i>  <b>1. Score integral</b>  <b>2. Nível de gravidade</b>	Complexidade fisiológica, com base em combinações de diagnósticos, sinais e sintomas específicos de doenças  Scores $\geq 0$  Nível de 1 a 4	Dados clínicos específicos de cada categoria. Ambas versões têm resultados da admissão e resultado máximo.  Soma do quadrado dos resultados para as variáveis clínicas individuais.
<i>MedisGroups (Atlas MQ)</i>  <b>1. Versão original</b>  <b>2. Versão empírica</b>	Instabilidade clínica indicada pela mortalidade intra-hospitalar (nível 0 a 4)  Probabilidade de morte intra-hospitalar (0-1.0)	Ambas versões têm scores de admissão e internamento. Os mesmos elementos clínicos são recolhidos em todos os registos médicos.

**Fonte:** Hughes et al. How severity measures rate hospitalized patient, 1996

As medidas de gravidade diferem quanto a origem dos dados, definição da gravidade, propósito da classificação e scores e intervalos de tempo durante a hospitalização em que os dados são extraídos. A fonte dos dados é a mais importante distinção prática entre as medidas de gravidade (Hughes et al, 1996).

A recolha de dados pode ser a partir de registos médicos como *MedisGroups*, APACHE II e CSI ou de resumos de alta como AIM, PMCs APR-DRGs, Índice de Comorbilidades, *Disease Staging* e R-DRGs (Costa, 1991).

#### **1.1.4. Dimensões do risco**

Nenhum método de ajustamento pelo risco pode contabilizar todas as dimensões relevantes, pelo menos por agora.

Muitos dos métodos de ajustamento pelo risco existentes baseiam-se nos dados administrativos. A operacionalização é feita através de grupos de algoritmos de códigos de diagnósticos da Classificação Internacional de Doenças 9ª Revisão Modificação Clínica (CID-9-MC) em categorias clinicamente significativas. As variáveis podem ser inseridas agrupadas no CID-9-MC nos seus próprios modelos, permitindo que cada variável da condição encontre seu próprio peso nos seus conjuntos de dados (Iezzoni LI, 2012).

Na identificação dos candidatos a dimensões de risco são utilizadas várias estratégias e desenvolvidas hipóteses sobre como cada um se relaciona com os resultados de interesse. Pode ser através da ajuda de relatórios publicados de ensaios randomizados e outros estudos clínicos. A outra estratégia é perguntar a peritos médicos o que é fundamental para envolver os clínicos no desenvolvimento dos ajustadores pelo risco e para a promoção da credibilidade clínica (Iezzoni, 1994).

#### 1.1.4.1. Categorias de dimensões do risco

Segundo Iezzoni (1997, 2003 e 2012; Knaus *et al.*, 1985 e 1991; Horn, 1988; Le Gall, Lemeshow e Saulnier, 1993; Lemeshow *et al.*, 1993), as dimensões a considerar no ajustamento pelo risco são classificadas em seis categorias principais:

- Genética (traços geneticamente determinados, predisposição genética para condições de saúde e para comportamentos relacionados com a saúde);
- Características demográficas (sexo, idade e etnia);
- Fatores clínicos (estabilidade fisiológica aguda, clínica aguda, diagnósticos e outras condições de saúde, extensão e gravidade dos diagnósticos, funções sensoriais, estado funcional, estado cognitivo, saúde mental e emocional);
- Fatores socioeconômicos (características e composição do agregado familiar, educação, estado civil, orientação sexual, recursos econômicos, ocupação e emprego, habitação, crenças e comportamentos culturais);
- Comportamentos e atividades relacionadas com a saúde (utilização de álcool e tabaco, de drogas ilícitas; práticas sexuais seguras; dieta e nutrição; excesso de peso e obesidade);
- Qualidade de vida, atitudes e percepções (preferências e expectativas pelos serviços de cuidados de saúde, comportamentos e crenças religiosas, estado de saúde e qualidade de vida, escolha de cuidados de fim de vida).

Descreve-se de seguida de acordo com a relevância para o estudo as dimensões do risco.

#### **Genética**

Estudos de mapeamento das funções dos genes e dos seus papéis nas doenças e consequências das interações no ser humano (Guttmacher and Collins, 2002).

## **Idade**

Na grande maioria das situações clínicas, os idosos apresentam maiores probabilidades de desenvolverem situações adversas do que os jovens (Iezzoni, 1997b). A idade é um fator de risco independente das outras características dos doentes, documentada no APACHE, MPM, SAPS (Wagner, Knaus e Draper, 1983). O APACHE III, na previsão da mortalidade hospitalar em doentes de UCIs atribui pontos para idade, de 0 para doentes com idades inferiores 45, e 24 pontos, para doentes com mais de 85 anos (Knaus et al, 1991).

## **Género**

As diferenças entre os sexos podem constituir um fator adicional a considerar na avaliação do risco de morte devido as diferenças anatómicas, fisiológicas e hormonais existentes (Iezzoni, 1997d), embora a grande maioria dos sistemas desenvolvidos não inclui o sexo como fator de risco de morte (Iezzoni, 1997b).

## **Raça**

Raça e etnia são difíceis de estabelecer de forma clara, sólida e significativa. As definições e as construções operacionais variam bastante (Barton e Heckscher, 2002).

## **Estabilidade fisiológica aguda**

Abrange o funcionamento fisiológico dos doentes no momento, comprovado pelas medidas homeostáticas básicas, como sinais vitais, eletrólitos séricos, achados hematológicos, oxigenação arterial e funcionamento neurológico. O objetivo é avaliar se os doentes estão em iminente risco de morte (Iezzoni, 1997d).

Os modelos de ajustamento pelo risco de doentes graves das UCIs têm indicadores de função fisiológica aguda nos seus cerne, como comprovam o APACHE II (Knaus et al, 1985), o SAPS II (Le Gall, Lemeshow e Saulnier 1993), o MPM (Lemeshow et al., 1993 e 1994) e o Pediatric Risk of Mortality (PRISM) (Pollack, Ruttimann e Getson, 1998).

Estes parâmetros fisiológicos são medidos de forma rotineira e servem de orientações aos médicos que tratam os doentes. Os dados obtidos produzem um score fisiológico agudo, credível, extraído dos programas no computador (Junger et al, 2002).

As medidas de estabilidade clínica aguda são usadas para prever o risco de morte iminente em doentes graves, complementados no MPM com a presença de cancro metastático, insuficiência renal crónica e cirrose (Lameshow et al., 1993). O SAPS II incorpora idade e cancro metastático, neoplasia hematológica e Síndrome de Imunodeficiência Adquirida (Le Gall, Lameshow e Saulnier 1993). O APACHE III acrescenta idade e pontuação para sete comorbilidade (Knaus et al 1991).

## **Diagnóstico**

Citação de Feinstein (1967) “diagnóstico, é o sinal central do pensamento no tratamento de um doente. O diagnóstico dá um nome à doença do doente, o pensamento passa cronologicamente para trás para decidir sobre patogénese e etiologia da doença. Do diagnóstico também, o pensamento passa cronologicamente para prever o prognóstico e escolher o tratamento. Como linguagem de comunicação clínica, os rótulos de diagnóstico transmitem uma compreensão fácil do conteúdo do pacote ...”.

Quando o diagnóstico é usado como dimensão de risco, ter um diagnóstico definitivo que atende a padrões exigentes pode não ser fundamental. No entanto, diagnósticos questionáveis (não confirmados) podem afetar a utilidade do diagnóstico como uma dimensão de risco (Iezzoni, 2012).

## **Gravidade e extensão dos diagnósticos**

O conceito de gravidade tem muitos estratos, geralmente organizadas à volta de prognósticos ou expectativas acerca dos doentes, cujo resultado clínico depende da extensão e natureza das doenças. A determinação dos níveis de gravidade dentro da categoria de diagnóstico muitas vezes é importante (Gonnella e Louis, 2005).

Na abordagem do doente, é também igualmente importante o estabelecimento do prognóstico na determinação do processo de tratamento do doente. (Gonnella, Hornbrook e Louis, 1984).

O termo prognóstico incita a mesma pergunta como a palavra risco: prognóstico de qual resultado, em relação a qual período? Dependente do diagnóstico, da extensão e do resultado de interesse, o prognóstico de uma situação clínica pode ser perspectivado a curto, a médio e



longo prazo (Hughes *et al.*, 1996), e é frequentemente considerado em termos de morte (Blumberg, 1987).

## **Comorbilidades**

Comorbilidades correspondem a existência de outros diagnósticos não relacionados com o diagnóstico principal (Iezzoni, 1997b). Geralmente estão presentes na admissão no hospital e são frequentemente considerados indicadores válidos na previsibilidade de morte a médio e longo prazo (Charlson *et al.*, 1987) e a curto prazo (Knaus *et al.*, 1981 e 1985; Le Gall, Lemeshow e Saulnier, 1993; Lemeshow *et al.*, 1993).

A comorbilidade dependendo do contexto, pode ser aguda (por exemplo, ataque cardíaca durante a admissão para tratamento de cancro do colon). A distinção entre comorbilidades agudas, complicações, doença iatrogénica, eventos adversos muitas vezes não é fácil. O momento da ocorrência do evento na fonte de dados e a confiabilidade são fundamentais. A possibilidade de acrescentar presente na admissão na folha de registo dos diagnósticos da nota de alta ajuda a ultrapassar esta dificuldade e auxilia na distinção de novas condições que ocorrem durante o internamento (Iezzoni 2012).

A avaliação das comorbilidades para inclusão no ajustamento pelo risco pode ser feita com (1) atribuição de um único score que representa a combinação de efeito de todas as condições de saúde por exemplo, *Physical Status Classification of the American Society of Anesthesiologists* (ASA) (Cohen *et al.* 2009; Hightower *et al.* 2010); (2) adicionando pontos individuais as condições de saúde para gerar um único score por exemplo o, Índice de Comorbilidade de Charlson; (3) através da introdução de variáveis de cada condição em modelos de regressão logística e permitir que cada um gere o seu próprio peso no conjunto de dados específicos, por exemplo, os investigadores do Research AND Development (RAND), (Keeler *et al.* 1990).

Elixhauser e colaboradores (1998) identificaram 30 comorbilidades utilizando dados de resumos de alta codificados com repercussão na duração de internamento, custos e mortalidade hospitalar.

## **Estado funcional**

O estado funcional basal é um preditor robusto de subseqüentes estados funcionais e de outros resultados, tal como morte iminente, consumo de recursos e satisfação com os cuidados, mas muitas vezes não está registado nos resumos de alta. Por exemplo nos doentes com pneumonia as dificuldades presentes na marcha predizem significativamente a mortalidade 30 dias após a hospitalização. (Daley et al, 1988; Iezzoni 1997b).

## **Outras dimensões de risco**

As restantes dimensões do risco — a situação psicológica e cognitiva dos doentes, os atributos culturais, étnicos e sócio-económicos dos doentes e as atitudes e preferências dos consumidores — são igualmente consideradas relevantes no apuramento da probabilidade de morte (Holmes et al. 2009; Baker et al., 2002b; Hanchate et al. 2009; Phelps et al. 2009; Danis et al. 1994; Iezzoni 2012)

Contudo, limitações de natureza prática, essencialmente as relacionadas com a dificuldade e fiabilidade na recolha dos dados, têm conduzido a que estas últimas dimensões do risco não sejam muito utilizadas nos sistemas de ajustamento pelo risco.

### **1.1.5. Suporte de recolha de dados**

A recolha de dados para desenvolvimento de um modelo de ajustamento pelo risco depende da definição do suporte de recolha de informação e a da definição dos momentos de medição (Costa, 2005).

Na operacionalização dos modelos de ajustamento ao risco são utilizados dados administrativos ou dados clínicos (Iezzoni, 1997c).

Quanto a melhor fonte de recolha de informação não existe unanimidade. Alguns autores (Fetter et al., 1980; Gonnella, Hornbrook e Louis, 1984; Gonnella et al., 1990) defendem a utilização dos resumos informatizados de alta hospitalar, outros (Knaus *et al.*, 1981; Le Gall et al, 1993; Lemeshow *et al.*, 1993) defendem que os elementos dos resumos uniformizados de alta devem ser complementados com dados dos processos clínicos.

Por dados administrativos entende-se os dados contidos no resumo de alta, enquanto por dados clínicos entende-se, para além daqueles, também os dados constantes do processo clínico, essencialmente aqueles que permitem caracterizar a história e o exame objetivo do doente, bem como os resultados dos meios complementares de diagnóstico (Costa, 2005).

Atualmente, no que se refere à atividade hospitalar em internamento, em Portugal existe informação de rotina sob a forma de dados administrativos, recolhida através do resumo de alta e generalizada nos hospitais públicos. Do resumo de alta consta:

- Identificação do hospital;
- N° do processo clínico;
- Elementos para a caracterização genérica do doente (sexo e data de nascimento);
- Residência;
- Entidade financeira responsável pelo pagamento e nº de beneficiário;
- Tipo de admissão (urgente, programada);
- Descrição das transferências internas (códigos de serviço, com a respetiva data de entrada e saída – até 20);
- Destino após a alta (para o domicílio, para outra instituição com internamento, serviço domiciliário, saído contra parecer médico, falecido);
- Diagnóstico de admissão;
- Códigos de diagnóstico (codificados através da CID-9-MC até 20, com menção do principal);
- Códigos de causa externa de lesão / efeito adverso (codificados com CID-9-MC e até 20);
- Códigos de procedimento (codificados através da CID-9-MC e até 20);
- Médico responsável;
- Médico codificador;
- Informações aplicáveis apenas a situações específicas: hospital de proveniência; hospital de destino e motivo de transferência; código da morfologia tumoral; peso à nascença; data da 1ª intervenção cirúrgica; nº de dias de internamento em UCI.

A recolha de dados clínicos geralmente é demorada e dispendiosa, o que limita a sua generalização (Lawson et al, 2012; Lane-Fall MB, 2013).

Do ponto de vista do investigador ou do gestor hospitalar, o uso de dados administrativos é atraente porque são fáceis de colher e permitem grandes amostras, o que

confere poder estatístico. Do ponto de vista do clínico cujo cuidado é avaliado, ou na perspectiva do doente ou pagador para comparar prestadores, os dados administrativos podem ser problemáticos porque estão sujeitos à documentação e as práticas de codificação das instituições, que podem ou não refletir com precisão os cuidados clínicos (Iezzoni, 2013).

No caso da utilização dos resumos de alta a fiabilidade não é boa (basicamente informação sobre diagnóstico principal e diagnósticos secundários). Segundo Costa e Lopes (2004) as principais razões para a fiabilidade não ser boa está relacionada com o grau de subjetividade na escolha do diagnóstico principal e também devido à sobreposição de categorias existente no CID-9-MC.

Por outro lado, no caso da utilização dados clínicos a objetividade é maior, proporcionada pelos sinais vitais e resultados dos exames complementares de diagnóstico, em que praticamente os problemas que podem existir dizem respeito somente ou a perdas de informação ou a erros de transcrição, o que confere uma maior objetividade aos sistemas que recorrem a estes elementos (Iezzoni et al., 1995b e 1996d; Iezzoni, 1997b).

A maioria dos dados administrativos não distingue comorbidades de complicações ou entre doenças agudas e crónicas (Fokkema et al 2014); não incluem critérios diagnósticos e outras informações clínicas, como pressão arterial e resultados laboratoriais (Hannan et al, 1997; Jollis et al, 1993; ou subestimam gravidade. (Shwartz et al, 1996b).

Num estudo Bensley e colaboradores (2013) afirmam que os dados administrativos não são adequados para determinar o estado dos sintomas e estados de alto risco, devido à falta de especificidade com os códigos de diagnóstico CID-9-MC, uma vez que não conseguem fornecer informações sobre a gravidade e início temporal da doença.

Os dados administrativos têm outras limitações como não conter informação acerca das preferências dos doentes, tal como indicação de não reanimar e pedidos apenas de medidas de conforto. Estas lacunas, podem comprometer a utilidade dos dados administrativos para produzir informações significativas sobre as taxas de mortalidade (Iezzoni LI 2012).

Em compensação, vários estudos comprovam a precisão estatística dos dados administrativos quanto dos dados clínicos. Iezzoni e colaboradores (1995a e 1997c),

concluíram que os dados administrativos prediziam a mortalidade intra-hospitalar em doentes com EAM com a mesma precisão dos dados clínicos (Iezzoni et al, 1998). Romano (1994) e Pine e colaboradores (1997) encontraram as mesmas previsões de mortalidade utilizando tanto dados administrativos como dados clínicos.

Em certos contextos, o ajustamento pelo risco pode obrigar a combinação de dados administrativos com clínicos selecionadas, como na cirurgia cardíaca. (Guru et al, 2008).

#### **1.1.6. Análise da qualidade dos dados e intervenções corretivas**

Os melhores modelos de ajustamento pelo risco são produzidos da combinação do julgamento clínico com modelação empírica, e com grandes bases de dados.

Os autores dos modelos de ajustamento pelo risco devem escolher cuidadosamente as suas estratégias de modelação e técnicas de validação que, entretanto, variam de acordo com a finalidade do modelo (comparação de prestadores ou tratamentos) ou previsão de eventos futuros (Shahian et al., 2004).

No desenvolvimento dos modelos os dados antes de serem analisados devem ser “*limpos*” que consiste na verificação de alcance, identificação de ocorrências impossíveis, pesquisa de dados inválidos e descrição das frequências de dados em falta ou elementos de dados mal especificados. (Iezzoni, 2012).

A falta de dados pode ocorrer por várias razões e a decisão de como lidar com a situação deve ser ponderada e varia segundo os contextos. No APACHE os parâmetros em falta são considerados normais, na presunção de que se o não fosse teriam sido registados. Quando a falta de dados tem consequência, deve ser descortinada as razões da indisponibilidade (por exemplo doentes com cancro metastizado que recusam a realização de análises de sangue e optam apenas por medidas de conforto). (Blumberg 1991).

Por outro lado, também o excesso de dados causa problema, portanto deve ser resolvido e no caso de impossibilidade, minimizado. A incorporação de demasiados preditores

é potencialmente nefasta como mostrou um estudo de mortalidade hospitalar da MedisGroups (Iezzoni et al. 1992b).

As variáveis de qualidade não confiável ou suspeita devem ser removidas, utilizando a metodologia da examinação das associações bivariada entre preditores individuais e os resultados. Atualmente está aconselhada a utilização de abordagens intensivas em computador na identificação de variáveis importantes para inclusão nos modelos de regressão logística (Normand et al, 1996).

#### **1.1.7. Técnicas de modelação**

As dimensões de risco são inseridas nos modelos sob a forma de variável independente categórica ou contínua. A dimensão de risco, idade, pode ser tratado como uma variável contínua ou categórica em dois ou mais grupos. No APACHE III atribui pontos nas diferentes categorias de idades (Knaus et al. 1991).

Variáveis clínicas, como pressão arterial e temperatura, têm relações complexas com vários resultados. Os valores extremos da pressão arterial estão associados a maior probabilidade de morte iminente. (Daley et al, 2008). Na maioria das variáveis clínicas a relação com o resultado não é linear. Geralmente é usado o julgamento clínico combinado, com análises estatísticas na determinação da área de variação entre limites superiores e inferiores numa escala específica nas variáveis contínuas e a relação do resultado dentro destes intervalos (Iezzoni 2012).

Na construção dos modelos habitualmente são usados dois procedimentos: “*stepwise*”. No tipo “*forward stepwise*” é adicionado uma variável de cada vez. Em cada fase, a variável adicionada é aquela que contribui melhor para modelar naquele passo. No tipo “*backward*” o procedimento é iniciado com todas as variáveis no modelo e, em seguida, elimina as variáveis que contribuem menos para o ajustamento do modelo uma a uma. Cada técnica inclui opções de libertar as variáveis previamente adicionadas (seleção “*forward*”) ou adicionar variáveis anteriormente descartadas (seleção “*backward*”). Normalmente, a modelação é considerada completa quando as variáveis adicionadas ao modelo são estatisticamente significativas  $p < 0,05$  ou  $0,10$ .

O principal desafio no desenvolvimento dos modelos é identificar as interações importantes ou efeitos não aditivos, entre as variáveis preditoras. Knaus e colaboradores (1991) utilizaram resultados de regressão logística, mais julgamento clínico para estudar interações entre as variáveis fisiológicas no APACHE III, na avaliação da ponderação individual e combinada de variáveis.

#### **1.1.8. Validade e confiabilidade dos modelos de ajustamento pelo risco**

A validade e confiabilidade dos métodos de ajustamento pelo risco constituem uma necessidade, mas a demonstração é complexa e algumas vezes ilusório (Iezzoni 2012).

Segundo Donabedian (1980), a validade abrange dois grandes domínios. O primeiro tem a ver com a precisão dos dados e das medidas que são construídas com esses dados; segundo com a justificação das inferências que são extraídas dos dados e da medição.

A validade e a confiabilidade dos modelos de ajustamento pelo risco são condicionadas pelo avanço do conhecimento clínico e da mudança na prática terapêutica. Portanto, estes métodos precisam de atualizações periódicas de modo a refletir os novos conhecimentos clínicos, novos códigos de diagnósticos e procedimentos. (Iezzoni, 2012).

As dimensões da validade dos modelos de ajustamento pelo risco integram:

1. Validade facial, quando o método contém as variáveis que lhe permite fazer o que é pretendido fazer;
2. validade de conteúdo, quando método incorpora todos os fatores de risco que devem estar incluídos. Está dependente da informação da fonte dos dados, especialmente dos dados administrativos que contêm informações clínicas limitadas (Iezzoni LI, 2012);
3. Critério e validade de construção - critério de validade, esclarece em que extensão uma medida se correlaciona com a medida *gold standard*. Enquanto, a validade de construção ou validade de correlação verifica-se, quando é encontrada uma associação positiva e forte com outras medidas credíveis, como a avaliação do risco por peritos médicos;

4. Validade preditiva expõe quanto o ajustador do risco prediz um resultado. O desempenho do ajustador de risco derivado empiricamente, deve ser reconhecido no conjunto de dados usados originalmente e em dados de validação. Para testar a validade preditiva muitas vezes os casos são divididos em dois ou mais subgrupos, usando um para desenvolver o modelo preditivo e o outro (ou outros) para testá-lo. (Wong et al 2011; Escobar et al. 2008);
5. Validade de atribuição, reflete a dimensão em que os observadores podem identificar a causa real das variações dos resultados entre os doentes (por exemplo, o porquê da diferença entre as taxas de mortalidade). A validade de atribuição é difícil de medir (Iezzoni, 2012).

## **Confiabilidade**

Confiabilidade aborda uma questão básica: se um processo for repetido, por outra pessoa, seguindo regras idênticas, ocorrerão os mesmos resultados. As medidas podem ser replicadas pelo mesmo observador ou avaliador (confiabilidade intra-avaliador) ou por avaliadores diferentes (confiabilidade inter-avaliador).

A confiabilidade dos modelos baseados em dados administrativos depende da precisão e da extensão das informações extraídas do registo médico ou de outras fontes primárias, tais como da consistência dos códigos de diagnóstico e dos procedimentos atribuídos. A confiabilidade dos ajustadores de risco baseados em dados administrativos, não foi testada diretamente, mas inferida a partir de estudos de confiabilidade de codificação (Iezzoni, 2012).

No caso dos ajustadores de risco baseados em dados clínicos, quando os dados são extraídos de fonte eletrónica geralmente são mais confiáveis que os dados extraídos manualmente. A confiabilidade depende do número de dados, da complexidade e da subjetividade dos julgamentos necessários para identificar variáveis, da qualidade dos registos e de outros fatores (Iezzoni, 2012).

Os métodos que classificam os casos com base num único código de diagnóstico são menos confiáveis (uma única inconsistência de codificação poderá afetar significativamente a classificação de risco) do que aqueles que dependem de múltiplos códigos (Iezzoni, 2012).



O *Kappa* é a medida mais comum da confiabilidade dos instrumentos com escalas categóricas e ordinais (Cohen 1960).

## **1.2. Sistemas de classificação de doentes**

### **1.2.1. Sistemas genéricos de classificação de doentes**

#### **1.2.1.1. Diagnosis Related Groups**

Ao longo dos anos foram desenvolvidos vários sistemas de classificação de doentes, genéricos e específicos como os Diagnosis Related Groups (DRG) ou Acuity Index Method (AIM) que privilegiam a utilização de recursos, outros a gravidade como Disease Staging, All Patient Refined DRGs (APR-DRG), *Acute Physiology and Chronic Health Evaluation* (APACHE), *Simplified Acute Physiology Score* (SAPS), *Mortality Probability Models* (MPM) e ainda outros que fundem os dois princípios, como os Patient Management Categories (PMCs). (Iezzoni, 1997a).

Os DRG desenvolvidos em 1970 na Universidade de Yale, com o objetivo de implementar um sistema de classificação de doentes que pudesse estar na base de um modelo de pagamento prospetivo (Mateus, 2010; Bentes, 1998; Bentes et al., 1996; Urbano e Bentes, 1990). Agrupam os episódios de internamento em categorias homogêneas quanto ao consumo de recursos, considerando-se para esse efeito a duração do internamento (Fetter et al., 1980).

O DRG foi desenvolvido a partir da análise estatística de uma base de dados contendo informação demográfica e clínica de cerca de 500 000 episódios, complementada com avaliação clínica para assegurar que as categorias formadas por critérios estatísticos eram clinicamente coerentes (Fetter et al., 1980).

A atribuição de um DRG a um episódio de internamento é realizada a partir da informação registada no resumo de alta tendo por base o diagnóstico principal, diagnósticos secundários, procedimentos cirúrgicos, idade e destino após alta.

O DRG é estruturado em conjuntos de categorias designadas Grandes Categorias Diagnósticas (GCD), que correspondem, em regra, a sistemas orgânicos. Cada GCD contém

um conjunto de DRG, que podem ser cirúrgicos ou médicos, consoante dizem respeito ou não a um procedimento realizado no bloco operatório (Hornbrook, 1982). Acrescenta-se que existem DRG quer médicos quer cirúrgicos onde são agrupados os doentes com registo de diagnósticos secundários que evidenciaram ter influência na duração de internamento (cuja designação inclui a expressão «CC» — complicações ou comorbilidades). Na elaboração do GDH ainda foi criada (1) DRG para determinados grupos etários porque a idade permitiu a redução da variância da duração de internamento dentro das categorias criadas e, (2) em certos casos, para destinos após alta específicos (por exemplo, a morte).

A codificação sistemática, em Portugal, dos episódios de internamento em hospitais de agudos com CID-9-CM iniciou-se em 1 de janeiro de 1989, correspondendo à introdução dos DRGs (Portaria n.º 325/89 de 4 de Maio).

A fonte de recolha de dados que serve de base aos diferentes tipos de DRG tem-se mantido inalterada. A diferença reside na dimensão e nos resultados alcançados no uso dessa informação para formar grupos coerentes do ponto de vista clínico e estatístico (Mullin et al, 2002).

Além do financiamento o DRG têm outras funções, como (1) avaliar a qualidade dos cuidados; (2) analisar de protocolos de tratamento, relacionados com as condições clínicas ou distribuição demográfica, (3) “*benchmarking*” e a análise de resultado obtidos utilizando a estrutura clínica do DRG (Urbano e Bentes, 1990; Bentes, 1996; Mateus, 2000; Santana, 2005; Averill et al, 2008); (4) auxiliar na avaliação da utilização dos recursos pelos serviços, o que permite aos hospitais avaliar e gerir os custos (Urbano e Bentes, 1990).

Os DRG estão capacitados para medir seis dimensões do casemix (Averill RF, 2008):

- Gravidade da doença - extensão da descompensação fisiológica ou perda de função do sistema de órgãos;
- Risco de Mortalidade - probabilidade de morrer;

- Prognóstico - provável resultado de uma doença, incluindo a probabilidade de melhoria ou deterioração da gravidade da doença, a probabilidade de recorrência e o provável tempo de vida;
- Dificuldade de tratamento - problemas de manuseio do doente que uma determinada doença coloca ao profissional de saúde. Esses problemas de manuseio estão associados a doenças sem um padrão claro de sintomas, doenças que exigem procedimentos sofisticados e tecnicamente difíceis, e doenças que exigem monitorização e supervisão rigorosos;
- Necessidade de intervenção - consequências em termos de gravidade da doença que a falta de atendimento imediato ou contínuo produziria;
- Intensidade do recurso - volume relativo e os tipos de serviços diagnósticos, terapêuticos e de camas usados no manejo de uma doença específica.

Os DRG ao longo dos anos, foram sujeitos a aperfeiçoamentos e como resultado estão disponíveis as seguintes versões: *Medicare DRGs*, *Refined DRG (R-DRG)*, *All Patient DRG (AP-DRG)*, *Severity DRG (S-DRG)*, *All Patient Refined DRG (APR-DRG)* e *International Refined DRG (IRDRGs)* (Averill et al., 1998; Mullin et al, 2002).

Os DRG mais atuais além da preocupação com o consumo de recursos estão também focados na gravidade da doença, na tentativa de explicar as variações no consumo de recursos dentro de um DRG. As novas versões também permitem comparar hospitais, implementar e avaliar a qualidade dos cuidados e servir de suporte à gestão interna, ao planejamento, entre outras funções (Averill et al, 2013).

A seguir são descritas as versões atuais dos DRG com maior aprofundidade o APR-DRG que faz parte do estudo.

**Medicare DRG** – foca-se essencialmente na população idosa abrangida pelo Medicare. A sua aplicabilidade é na classificação e financiamento de episódios de internamento. Todavia, esta modalidade foi também usada na classificação e financiamento de episódios de internamento de recém-nascidos, crianças e adultos não idosos (Muldoon, 1999).

**Refined DRG (RDRG)** - emergiu da revisão dos CCs dos DRG da *Medicare* (Freeman, et al, 1995), na presunção que as CCs aumentariam a utilização de recursos. Foram atribuídos níveis de complexidade a todos os diagnósticos secundários considerados CC nos DRG *Medicare*. Nos doentes cirúrgicos 4 níveis de complexidade (não-CC, CC moderada, CC major e CC catastrófica) e aos doentes médicos três níveis de complexidade (não-CC, moderada ou major CC e CC catastrófica) (Averill et, al, 1998).

**Severity DRG (SDRG)** - desenvolvido (1993) da reavaliação nos CCs dos DRGs da *Medicare*, pela HCFA, com exclusão dos DRG associados a gravidez, recém-nascidos e doentes pediátricos. Todas as divisões CCs nos DRG da *Medicare* foram eliminadas e foi fundida o GCD, 24 da *Medicare*. Nos SDRG, um doente é atribuído a um subgrupo correspondente ao diagnóstico secundário de mais alto nível (Averill, et al, 1998).

**Internacional Refined DRG (IR- DRG)** - desenvolvido, para suprir as necessidades da comunidade internacional em produzir o seu próprio software. Permite um determinado país localizar os IR -DRG usando os seus próprios códigos de diagnóstico e de procedimentos. Os IR - DRG usam a severidade ajustada para descrever a exigência de recursos com base nas características individuais dos doentes (Mullin, et al, 2002). (3M HIS).

**All Patient DRG (AP-DRG)** - a conceção (1988) incluiu bases de dados não só dos beneficiários do *Medicare*, como também outros episódios de internamento (Averill et al., 1998). No desenvolvimento, o New York State Department of Health (NYDH) e 3M Health Information Systems (3M HIS) submeteram o *Medicare* DRG a um conjunto de alterações que incluíram a reestruturação da GCD neonatal e a criação de uma GCD para o Vírus da Imunodeficiência Humana (VIH). Houve ainda acréscimo da GCD 25, de doentes com traumas múltiplos, de alterações nos transplantes, doentes sob ventilação mecânica de longa duração, fibrose quística, distúrbios nutricionais, cuidados obstétricos de alto risco, leucemia aguda, hemofilia e anemia falciforme (Averill et al., 1998).

Na conceção dos AP-DRG foram revistas e identificadas complicações e comorbilidades que, pelo elevado impacto sobre o consumo de recursos são consideradas major. No seguimento, foram acrescentadas em cada GCD, DRG, onde são agrupados os doentes com complicações e comorbilidades (Averill et al., 1998).

**All Patient Refined DRG (APR- DRG)** - desenvolvido (1999) pela 3M HIS e *National Association of Children`s Hospitals and Related Institutions* (NACHRI) a partir dos AP- DRG com exceção da GCD base do recém-nascido. No APR- DRG foi eliminada a idade e CCs e, as CCs major substituídas por dois conjuntos de quatro subgrupos de doentes que abordam diferenças relacionadas com níveis de severidade da doença e de risco de mortalidade.

O APR-DRG comporta três descritores distintos: APR- DRG base; subgrupo de severidade da doença e subgrupo de risco de mortalidade, cada um com quatro níveis: menor (1), moderada (2), major (3) e extremo da doença (4). A atribuição a um doente de um destes quatro subgrupos tem em consideração não só os diagnósticos secundários, mas também a interação entre os diagnósticos secundários, a idade, o diagnóstico principal e a presença de certos procedimentos não realizados no bloco operatório.

A avaliação da severidade da doença ou risco de mortalidade de um doente é específico do APR- DRG em que o doente está atribuído e a determinação da severidade da doença e do risco de mortalidade são específicos da doença. Os doentes com múltiplas doenças envolvendo vários sistemas de órgãos, apresentam severidade elevada, processo de doença mais extensa, e colocam dificuldades no tratamento e os resultados são piores.

O APR-DRG, conjuntamente com o subgrupo de severidade da doença, são utilizados para a avaliação da utilização de recursos ou o estabelecimento de protocolos de cuidados aos doentes. Os doentes com elevada severidade da doença, em geral, incorrem em elevados custos e têm risco mais elevado de mortalidade esperada e morrem mais.

Os dados usados pelo APR-DRG são: diagnóstico principal, diagnósticos secundários, procedimento principal, procedimentos secundários, idade, sexo, peso ao nascer, estado na alta e dias de ventilação mecânica.

Os APR-DRG classificam os doentes segundo o motivo de admissão, o que permite estabelecer relação clínica entre a admissão inicial e a readmissão (Averill et al, 2002).

Através do APR-DRG, os hospitais, os consumidores, os pagadores e os reguladores podem perceber os doentes que estão a tratar, os custos incorridos e, dentro de limites razoáveis, os resultados esperados dos serviços. Com o APR-DRG, é possível identificar áreas onde pode ser melhorada a eficiência e identificar problemas de qualidade.

Embora as subclasses estejam numeradas 1–4, os valores numéricos representam categorias e não pontuação. Por exemplo, a subclasse de severidade 4 em doentes com insuficiência cardíaca congestiva não é comparável à subclasse de severidade 4 dos doentes com fratura do membro inferior. Assim, não tem significado calcular os valores numéricos (1-4) de subclasses de severidade da doença ou risco de subclasses de mortalidade num grupo de doentes para calcular uma pontuação de severidade média. No entanto, as subclasses de severidade e risco de mortalidade de APR-DRG podem ser usadas para calcular um valor esperado de uma medida de interesse (por exemplo, tempo de permanência, custo, mortalidade), usando técnicas estatísticas, como taxa de padronização indireta (3M HIS).

Atualmente, os GDH além de servirem de base ao financiamento do internamento hospitalar, também são utilizados em contexto de ambulatório. (ACSS (Autoridade Central do Sistema de Saúde, 2015)

O agrupador atual de GDH é o *All Patients Refined* DRG, versão 31.0, que corresponde à versão da CID-9-CM de outubro de 2013 (Portaria 234/2015)

#### **1.2.1.2. Disease Staging**

O Staging é um sistema de classificação de doentes que pretende produzir grupos de doentes que requerem tratamentos idênticos e com resultados esperados idênticos (Garg et al., 1978; Gonnella, Hornbrook e Louis, 1984; Gonnella, Louis e Gozum, 1999).

Comporta duas modalidades: (1) uma clínica que classifica um episódio de internamento com base na informação recolhida relativamente aos exames objetivos, resultados laboratoriais e radiológicos e ainda de relatórios anátomo-patológicos e cirúrgicos (MEDSTAT, 2010) e (2) outra administrativa (*Coded*) desenvolvida da versão clínica, corresponde genericamente à classificação dos episódios de internamento a partir dos

diagnósticos codificados de acordo com CID-9-MC registrados no resumo de alta (Gonnella, Louis e Gozum, 2003).

A concepção do Staging iniciou com a identificação das doenças mais frequentes e mais relevantes no internamente em cada sistema orgânico e em cada etiologia. Depois, cada doença foi dada a peritos para definição estádios de evolução específica dessa doença. Esses estádios deveriam acarretar uma morbilidade/mortalidade significativamente diferente e ser clinicamente distinguíveis, bem como conter os sub-estádios considerados oportunos. Os estádios foram definidos de 1 a 4 e representam extensão da progressão da doença. Em algumas doenças é considerado o estágio 0, como é o caso de fatores de risco para desenvolvimento de doenças (história/antecedentes de neoplasia da mama). (Gonnella, Louis e Gozum, 1999). Na definição dos estádios, foi considerado apenas a história natural da doença e não foi tido em consideração o tipo de tratamento, (médico ou cirúrgico) (Gonnella, Hornbrook e Louis, 1984).

A etiologia e o estágio da doença determinam o risco de morte que aumenta com cada nível de estágio superior, podendo variar drasticamente do estágio de uma doença para outra doença (Gonnella e Louis, 2005). A etiologia da doença deve ser especificada ou a mais provável, sempre que possível. Se tal não acontecer, os doentes podem ser classificados de forma inadequada para efeitos de análise da utilização dos recursos, avaliação da qualidade dos cuidados e reembolso (Louis et al, 2004).

As características dos doentes, como estádios mais elevados da doença, idade elevada e doentes admitidos de urgência são mais propensas a ter níveis de gravidade e taxas de mortalidade mais elevadas. Todos estão associados a custos de tratamento mais elevados e maior tempo de permanência no hospital. Também os casos cirúrgicos têm permanências mais longas do que os casos não cirúrgicos (Gonnella et al., 1984).

O Staging permite (1) avaliar qualidade dos cuidados através da adequação da admissão hospitalar detetando potenciais problemas existente nos cuidados do ambulatório (Gonnella e Louis, 1988; Gonnella et al., 1990; Gonnella e Zeleznik, 1988; Manzoli, 2004) e (2) atua ao nível dos processos na elaboração de protocolos clínicos e ajuda na orientação de processo de cuidados e na avaliação dos resultados clínicos. (Gonnella e Louis, 1988; Gonn e 1995).

O *Staging* tem ainda outras aplicações, como nos ensaios clínicos (Markson et al., 1991); no planeamento de profissionais e de equipamentos de saúde (Conklin et al., 1984; Gonnella, Hornbrook, e Louis, 1983); na certificação de especialidade e de competências clínicas (Kaga e Gonnella, 1990); na educação médica (Rattner et al., 2001).

A codificação (Medstat, 2010) é efetivada com atribuição de categorias de doenças e estádios aos diagnósticos contidos nos resumos de alta. A cada admissão é atribuída uma doença principal, baseada no diagnóstico principal. O Disease Staging considera também os diagnósticos secundários registados e pode, quando se justifique, atribuir outras doenças além da principal (comorbilidades). Na identificação da doença principal e comorbilidades e os respetivos estádios é usada a seguinte metodologia (Gonnella, Louis e Gozum, 1999):

- Quando o diagnóstico secundário é considerado como uma complicação do diagnóstico principal, este altera o estágio do diagnóstico principal (por exemplo, quando a hipertensão arterial é o diagnóstico principal e se está na presença de retinopatia hipertensiva e de insuficiência renal como diagnósticos secundários, então estes últimos são considerados complicações do diagnóstico principal e o software identificará uma doença principal e um estágio e sub-estágio específico);
- Quando o diagnóstico secundário não está associado com o diagnóstico principal, é então definida uma comorbilidade, assim como o respetivo estágio e sub-estágio;
- Quando os diagnósticos secundários são sintomas, então nenhuma comorbilidade é considerada e o estágio da doença principal mantém-se.

A metodologia do *Staging* desenvolvida (Grupo MEDSTAT), centra-se na especificidade da doença e avalia a gravidade da doença através da determinação dos estádios da doença, expressos em níveis ordinais, portanto não são passíveis de serem calculados através da média das doenças na descrição da gravidade global da doença.

MEDSTAT: disponibiliza escalas preditivas, obtidas através de regressão múltipla de base de dados  $\pm$  15 milhões de registos de doentes. As regressões aplicadas para obtenção das escalas de custos e duração de internamento, foram executadas separadamente para cada DRG e DXCAT, porque a diminuição das taxas médias de custos no internamento não é linear. A intensidade do tratamento é mais elevada no início do internamento.



MEDSTAT 2001: na versão automatizada do Staging disponibiliza identificação dos estádios e sub-estádios da doença principal e comorbilidades, previsões por doente para a duração de internamento, para os custos, para a mortalidade, para as complicações e para as readmissões. Os valores esperados por doente são atribuídos em função da sua doença principal e estágio, idade, sexo, comorbilidades e estádios e se se trata de uma admissão urgente, considerando o seu impacto esperado sobre o resultado em análise.

As escalas preditivas de mortalidade do MEDSTAT foram produzidas da mesma base de dados a partir de todos os DRGs e/ou DXCATs, separando os DRGs médicos dos cirúrgicos. As taxas de mortalidade esperadas foram calculadas em quatro classes segundo o número de DXCAT ou DXCAT/DRG (< 300 ou > 300 altas).

A escala de mortalidade é calculada dividindo-se a mortalidade previsível, obtida de uma das quatro classes descritas em cima, pela taxa global de mortalidade hospitalar a partir de base de dados de desenvolvimento 100 vezes.

Medstat (2010): também disponibiliza um software que integra 37 complicações de cuidados que identifica se o doente estava em risco para uma determinada complicação e se a complicação estava no registo do doente. As variáveis de *inputs* no *software* do *Disease Staging* necessárias para identificar as complicações incluem: (1) códigos de diagnóstico principal e secundário do CID-9-MC (até 15 códigos); (2) códigos de procedimento do CID-9-CM; (3) Sexo; (4) Idade; (5) estado na alta; (6) duração de internamento.

### **1.2.2. Sistemas de classificação de doentes clínicos**

Os modelos clínicos de ajustamento pelo risco utilizados nas UCIs, permitem determinar a gravidade da doença e grau de disfunção de órgãos, avaliar o consumo de recursos e prever o resultado (geralmente resultado à alta hospitalar) embora possam fazer outras medições de resultados (por exemplo, sobrevivência aos 28 dias, pós a alta hospitalar) (Lemeshow, 1994). A combinação dos modelos fornece indicação mais precisa da gravidade da doença e do prognóstico (Vincent e Moreno, 2010).

Para desenvolver um sistema de score ideal, é indispensável, uma grande base de dados de doentes de muitas UCIs e de diferentes países. As variáveis utilizadas são agrupadas em cinco categorias: idade, alterações fisiológicas, comorbidades, diagnóstico agudo (motivo de admissão) e intervenções (Bouch e Thompson, 2008).

Os modelos clínicos podem ser gerais ou direcionados a um órgão ou sistema. Os gerais avaliam a gravidade da doença a quando da admissão e permitem prever o resultado [por exemplo, *Acute Physiology and Chronic Health Evaluation* (APACHE), *Simplified Acute Physiology Score* (SAPS), *Mortality Probability Model* (MPM)], os sistemas de score que avaliam a presença e a gravidade da disfunção orgânica [(por exemplo, *Multiple Organ Dysfunction Score* (MODS), *Sequential Organ Failure Assessment* (SOFA)].

**TABELA 2 INFORMAÇÃO SOBRE OS MODELOS CLÍNICOS**

Características	APACHE I	SAPS I	APACHE II	MPM <sup>a</sup>	APACHE III	SAPS II	MPM II <sup>b</sup>	SAPS 3	APACHE IV	MPM III
Ano	1981	1984	1985	1985	1991	1993	1993	2005	2006	2007
Países participantes	1	1	1	1	1	12	12	35	1	1
UCI participantes	2	8	13	1	40	137	140	303	104	135
Nº de doentes	705	679	5.815	2.783	17.440	12.997	19.124	16.784	110.558	124.855
Seleção das variáveis e dos seus pesos	PdP	PdP	PdP	RLM	RLM	RLM	RLM	RLM	RLM	RLM
Variáveis										
Idade	Não	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim
Origem do doente	Não	Não	Não	Não	Sim	Não	Não	Sim	Sim	Não
Estado cirúrgico	Não	Não	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim
Estado de saúde crónico	Sim	Não	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim
Fisiologia	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim
Diagnóstico agudo	Não	Não	Sim <sup>c</sup>	Não	Sim <sup>d</sup>	Não	Sim	Sim	Sim	Sim
Número de variáveis	34	14	17	11	26	17	15	20	142	16 <sup>e</sup>
Pontuação	Sim	Sim	Sim	Não	Sim	Sim	Não	Sim	Sim	Não
Equação preditiva da mortalidade	Não	Não	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim

APACHE, *Acute Physiology and Chronic Health Evaluation*; SAPS, *Simplified Acute Physiology Score*; MPM, *Mortality Probability Models*.

**PdP**: painel de peritos; **RLM**: regressão logística múltipla

<sup>a</sup>: Estes modelos são baseados em versões prévias, desenvolvidos pelos mesmos autores.

<sup>b</sup>: Os números apresentados referem-se ao componente de admissão do modelo (MPMII<sub>0</sub>). O MPM II<sub>24</sub> foi desenvolvido com base em dados de 15.925 doentes.

<sup>c</sup>: Escolhido de uma lista de 50 diagnósticos.

<sup>d</sup>: Escolhido de uma lista de 78 diagnósticos.

<sup>e</sup>: O MPM II<sub>24</sub> utiliza apenas 13 variáveis.

**Fonte:** Gestão e organização em Medicina Intensiva – Rui Moreno, 2000

### 1.2.2.1. APACHE

O APACHE foi o primeiro índice de gravidade geral desenvolvido e tem aplicabilidade na maioria dos doentes graves (Knaus et al, 1981).

O APACHE II é uma versão simplificada lançada em 1985 (Knaus et al) a partir de bases de dados de doentes de hospitais dos EUA e integra 17 variáveis sendo 12 fisiológicas. No seu desenvolvimento foram excluídos doentes com idade inferior a 18 anos, doentes queimados, doentes coronários e doentes no pós-operatório de cirurgia cardíaca. São registados os piores valores de cada variável fisiológica (com pontuação de 0 a 4) das primeiras 24 horas de internamento na UCI cujo somatório produz o *Acute Physiology Score* (APS). Do registo fazem parte ainda a idade, cirurgia de urgente/não urgente ou não cirúrgico, o estado prévio de saúde e o diagnóstico agudo (motivo de admissão). Esta última variável tem peso importante no cálculo da mortalidade previsível do doente, mesmo quando o estado de saúde prévio e o grau da disfunção fisiológica aguda sejam semelhantes. O sistema evidenciou boa calibração e bom valor discriminatório (Knaus et al, 1985).

O APACHE III lançado em 1991 (Knaus et al) e aperfeiçoado 1998. Contém mais variáveis que o APACHE II e os cálculos são feitos também com os piores valores das 1<sup>as</sup> 24 horas de internamento (pontuação de 0 a 4). A conversão da pontuação em probabilidade de mortalidade é feita através da utilização de equações logísticas individuais para cada um dos 78 diagnósticos e de 9 origens dos doentes. No seu desenvolvimento foram excluídos doentes com permanência inferior a 4 horas, menores de 16 anos e os doentes admitidos queimados,

doentes coronários e doentes pós cirurgia de *bypass* coronário. Prevê resultado de doente individual, identifica os fatores que influenciam as variações de resultados dos cuidados e tempo de permanência nas UCIs (Knaus et al, 1993). O seu desempenho é ligeiramente melhor do que o APACHE II.

A última versão disponível, APACHE IV foi desenvolvido em 2006 (Zimmerman et al). Fornece indicações de previsibilidade do tempo de internamento e valores de referência na avaliação e comparação de eficiência e de utilização de recursos nas UCIs.

#### **1.2.2.2. SAPS**

O SAPS foi desenvolvido em França em 1984 (Le Gall et al) como alternativa ao APACHE.

A versão SAPS II foi lançada em 1993 (Le Gall, Lemeshow e Saulnier), a partir de bases de dados de hospitais europeus e norte-americanos. Como no APACHE, a pontuação é calculada a partir dos piores valores das primeiras 24 horas de internamento. Também foram excluídos no desenvolvimento doentes com idade inferior a 18 anos, doentes queimados, doentes coronários, e pós cirurgia cardíaca. Inclui 17 variáveis sendo 12 fisiológicas mais a idade, o tipo de admissão e 3 variáveis relacionadas com doença subjacente. Foi validado numa grande base de dados (137 UCIs de 12 países). O *score* de 0 a 163 pontos. Não necessita de motivo de admissão para o cálculo da probabilidade de morte hospitalar.

O SAPS III, desenvolvido a partir de uma grande base de dados de 35 países, e de 7 regiões geográficas (Moreno et al, 2005). Inclui 20 variáveis divididas em três subescalas relacionadas com as características do doente antes da admissão, motivo da admissão e o grau de alteração fisiológica na 1ª hora (em contraste com a janela de 24 horas no SAPS II) antes e após a admissão na UCI. O *score* total varia de 0 a 217. Capacitado para avaliar a utilização dos recursos entre UCIs a partir de parâmetros padronizados dos recursos com base no tempo de internamento ajustados a gravidade da doença aguda (Rothen et al, 2007). O SAPS 3 mostrou boa discriminação e calibração (Moreno et al, 2005).

### **1.2.2.3. MPM**

O MPM tem vários modelos para prever mortalidade à admissão na UCI (MPM<sub>0</sub>) e às 24 horas (MPM<sub>24</sub>); para as 48 horas (MPM<sub>48</sub>) e 72 horas (MPM<sub>72</sub>) na UCI. O mais utilizado é MPM II, que foi desenvolvido em 1993 (Lemeshow et al) para doentes com permanência superior a 24 horas nas UCs. Compõe de dois modelos: MPM<sub>0</sub> com 15 variáveis da admissão e MPM<sub>24</sub> com 5 variáveis da admissão e 8 variáveis adicionais. Ao contrário do APACHE e SAPS onde as variáveis são ponderadas, no MPM II cada variável (exceto idade, que é inserida em anos), é designado como 1 ou 0 (presente ou ausente). A probabilidade de mortalidade é obtida através de uma equação de regressão logística e não como um score.

Os sistemas de score de disfunção de órgãos (SOFA; MODS e LODS) não foram projetados para prever a mortalidade na UCI, mas maior disfunção de órgãos (neurológico, cardiovascular, renal, respiratório, hematológico e hepático) está associada consistentemente a maior mortalidade (Metnitz et al, 2001)

### **1.3. Avaliação do desempenho dos modelos de ajustamento pelo risco**

Após o desenvolvimento de um método de ajustamento pelo risco, o passo seguinte é a avaliação do seu desempenho, que deverá ser realizado numa população diferente daquela usada para desenvolver o método (Le Gall, 2005), através da calibração e a discriminação do modelo (Lemeshow e Le Gall JR, 1994).

A calibração avalia o grau de correspondência entre a probabilidade estimada e a observada de mortalidade utilizando o teste estatístico de Hosmer-Lemeshow. Uma série de probabilidades de mortalidade esperada e observada, são comparadas e um valor de P derivado. A calibração é considerada boa se a mortalidade prevista estiver próxima da mortalidade observada (Ridley, 1998).

As razões de mortalidade observada/prevista são calculadas dividindo a taxa de mortalidade observada (ou seja, o número de mortes) pela taxa de mortalidade prevista (ou seja, a soma das probabilidades de mortalidade de todos os doentes analisados). Num modelo perfeitamente calibrado este valor deve ser um.

O score Z de Flora (1978) é baseado numa estatística que compara o número de sobreviventes observados na amostra a analisar com o número previsto de acordo com o modelo. A diferença é então padronizada e comparada com uma tabela da distribuição normal (Flora, 1978).

Os testes de Hosmer-Lemeshow são duas estatísticas de qui-quadrado propostas para a avaliação formal da calibração de modelos preditivos (Hosmer e Lemeshow, 2000). No teste H, os doentes são divididos em 10 grupos de acordo com as suas probabilidades: 0.0 a 0.1, 0.1 a 0.2... 0.9 a 1.0. Então, uma estatística qui-quadrado é utilizada e compara o número observado com o número previsto de mortes e o número observado e previsto de sobreviventes em cada um dos grupos.

A estatística resultante é então comparada com uma tabela de qui-quadrado com 8 graus de liberdade (no desenvolvimento do modelo) ou 10 graus de liberdade (na validação do modelo), no sentido de determinar se as discrepâncias observadas podem ser explicadas apenas com base na oscilação aleatória. Grandes diferenças sugerem que o modelo não está bem calibrado.

O teste C é semelhante, com os 10 grupos definidos por decis das probabilidades estimadas. Hosmer e Lemeshow demonstraram que o método de agrupamento na base da estatística-c é preferível sempre que muitas das probabilidades a analisar sejam pequenas (Hosmer e Lemeshow, 1989).

A discriminação analisa a capacidade do modelo para discriminar entre os doentes que morrem dos que sobrevivem, com base na mortalidade prevista. Os métodos incluem o cálculo da área sob a curva de ROC (receiver operating characteristic) ou usam uma matriz de classificação. As duas partes mais importantes da matriz de classificação são a sensibilidade e a especificidade (Ridley, 1998).

A curva de ROC é produzida por valores de sensibilidade-especificidade ao longo do intervalo de predição de mortalidade. A área sob a curva resultante representa o número de doentes que morreram. A curva é analisada usando processos estatísticos complexos computadorizados para avaliar a discriminação (Ridley, 1998). Se esta área for de 0,50, o

desempenho do método não é melhor do que uma moeda ao ar. Os desenvolvedores de modelos impõem que uma área sob a curva ROC seja  $> 0,70$  (Lemeshow, Le Gall, 1994).

A metodologia estatística, Hosmer-Lemeshow mais utilizada para avaliar a calibração de um modelo preditivo, pode ser influenciada por vários fatores, incluindo o número de co-variáveis a serem avaliadas, a maneira pela qual as observações com probabilidades iguais de resultado são classificadas e o tamanho da amostra (Krame e Zimmerman, 2007).

Depara-se com diversos problemas quando se pretende aplicar modelos de ajustamento pelo risco em populações diferentes ou a sub-grupos distintos das populações de onde o modelo foi originalmente construído (Iezzoni, 1999). Estes problemas foram designados por Flanders e colaboradores (1999) como o *caveat emptor*, que significa calibração imperfeita.

Justice, Covinsky e Berlin (1999) definiram precisão e generalização respeitante à utilização de modelos de ajustamento pelo risco, como descrito a seguir:

Na avaliação do desempenho dos modelos (Covinsky e Berlin 1999) devem ser considerados (1), a precisão (se as previsões coincidem com os resultados na amostra original de doente), avaliado por exame de calibração e discriminação; e a generalização (em que medida o modelo fornece previsões precisas em diferentes amostras de doentes).

Um sistema que não discrimina em estudos subsequentes provavelmente não estará calibrado (Harrell et al, 1996; Ash e Schwartz, 1994), mas muitos modelos perdem a calibração enquanto mantêm alguma discriminação. Como a importância relativa da calibração e discriminação dependem do pretendido, é importante testar ambos. Além disso, a generalização de um sistema preditivo depende da sua reprodutibilidade e transportabilidade. A reprodutibilidade pode ser avaliada usando técnicas de bootstrapping na amostra de desenvolvimento, mas a avaliação da transportabilidade exige novos dados.

Quanto maior forem os dados usados no desenvolvimento do modelo, maior é a sua capacidade para identificar as dimensões de risco importantes (variáveis preditores) e quantificar com precisão o risco

A generalização, também conhecida como validade externa (Ash e Shwartz, 1994; Knaus et al, 1985), é a capacidade de um sistema prognóstico em oferecer previsões precisas em amostras posteriores de doentes.

A validação externa do modelo de prognóstico é dificultada quando a sua utilização é diferente da que foi projetado. Por exemplo, Charlson e colaboradores (1987) desenvolveram o índice de comorbidade para prever a mortalidade de 1 ano, usando informações num coorte de 604 doentes internados no serviço de medicina de um Hospital da Nova York durante 1 mês, em 1984, e na sua validação usaram informações de 685 doentes tratados entre 1962 e 1969 por cancro da mama no Hospital Yale New Haven, o que é evidentemente muito tempo tendo em consideração a evolução dos procedimentos terapêuticos da neoplasia da mama e mudanças na população de doentes.

As técnicas de regressão logística multivariada, geralmente são utilizadas para se estabelecerem previsões sobre o resultado em análise (Iezzoni, 1997b), visto que permitem traduzir a relação existente entre uma variável dependente ou resultado e várias variáveis explicativas. Representam um método comum para a análise de relações em que a variável dependente não é contínua (Hosmer e Lemeshow, 1989). Esta é a situação mais frequente quando se pretende analisar a mortalidade hospitalar, em que a variável dependente é dicotómica, assumindo os valores 0 (alta vivo) e 1 (alta falecido) em função de um conjunto de variáveis independentes, frequentemente designadas por covariáveis (Hosmer e Lemeshow, 1989; Ruttiman, 1994).

Após o desenvolvimento de um método de ajustamento pelo risco, o passo seguinte é a avaliação do seu desempenho, que deverá ser realizado numa população diferente daquela usada para desenvolver o método (Le Gall, 2005), através da calibração e a discriminação do modelo (Lemeshow e Le Gall JR, 1994).

A calibração avalia o grau de correspondência entre a probabilidade estimada e a observada de mortalidade utilizando o teste estatístico de Hosmer-Lemeshow. Uma série de probabilidades de mortalidade esperada e observada, são comparadas e um valor de P derivado. A calibração é considerada boa se a mortalidade prevista estiver próxima da mortalidade observada (Ridley, 1998).



As razões de mortalidade observada/prevista são calculadas dividindo a taxa de mortalidade observada (ou seja, o número de mortes) pela taxa de mortalidade prevista (ou seja, a soma das probabilidades de mortalidade de todos os doentes analisados). Num modelo perfeitamente calibrado este valor deve ser um.

O score Z de Flora (1978) é baseado numa estatística que compara o número de sobreviventes observados na amostra a analisar com o número previsto de acordo com o modelo. A diferença é então padronizada e comparada com uma tabela da distribuição normal (Flora, 1978).

Os testes de Hosmer-Lemeshow são duas estatísticas de qui-quadrado propostas para a avaliação formal da calibração de modelos preditivos (Hosmer e Lemeshow, 2000). No teste H, os doentes são divididos em 10 grupos de acordo com as suas probabilidades: 0.0 a 0.1, 0.1 a 0.2... 0.9 a 1.0. Então, uma estatística qui-quadrado é utilizada e compara o número observado com o número previsto de mortes e o número observado e previsto de sobreviventes em cada um dos grupos.

A estatística resultante é então comparada com uma tabela de qui-quadrado com 8 graus de liberdade (no desenvolvimento do modelo) ou 10 graus de liberdade (na validação do modelo), no sentido de determinar se as discrepâncias observadas podem ser explicadas apenas com base na oscilação aleatória. Grandes diferenças sugerem que o modelo não está bem calibrado.

O teste C é semelhante, com os 10 grupos definidos por decis das probabilidades estimadas. Hosmer e Lemeshow demonstraram que o método de agrupamento na base da estatística-c é preferível sempre que muitas das probabilidades a analisar sejam pequenas (Hosmer e Lemeshow, 1989).

A discriminação analisa a capacidade do modelo para discriminar entre os doentes que morrem dos que sobrevivem, com base na mortalidade prevista. Os métodos incluem o cálculo da área sob a curva de ROC (receiver operating characteristic) ou usam uma matriz de classificação. As duas partes mais importantes da matriz de classificação são a sensibilidade e a especificidade (Ridley, 1998).

A curva de ROC é produzida por valores de sensibilidade-especificidade ao longo do intervalo de predição de mortalidade. A área sob a curva resultante representa o número de doentes que morreram. A curva é analisada usando processos estatísticos complexos computadorizados para avaliar a discriminação (Ridley, 1998). Se esta área for de 0,50, o desempenho do método não é melhor do que uma moeda ao ar. Os desenvolvedores de modelos impõem que uma área sob a curva ROC seja  $> 0,70$  (Lemeshow, Le Gall, 1994).

A metodologia estatística, Hosmer-Lemeshow mais utilizada para avaliar a calibração de um modelo preditivo, pode ser influenciada por vários fatores, incluindo o número de co-variáveis a serem avaliadas, a maneira pela qual as observações com probabilidades iguais de resultado são classificadas e o tamanho da amostra (Krame e Zimmerman, 2007).

Depara-se com diversos problemas quando se pretende aplicar modelos de ajustamento pelo risco em populações diferentes ou a sub-grupos distintos das populações de onde o modelo foi originalmente construído (Iezzoni, 1999). Estes problemas foram designados por Flanders e colaboradores (1999) como o *caveat emptor*, que significa calibração imperfeita.

Justice, Covinsky e Berlin (1999) definiram precisão e generalização respeitante à utilização de modelos de ajustamento pelo risco, como descrito a seguir:

**TABELA 3 CRITÉRIOS.** FONTE: JUSTICE ET AL, 1999

Termo	Critério ou definição
Precisão	Grau em que os resultados previstos correspondem aos resultados observados
Calibração	Probabilidade prevista não é nem muito alta nem muito baixa (geralmente exposta com curvas de calibração).
Discriminação	A classificação relativa do risco individual está ordenada corretamente (as taxas de eventos observados nas pessoas com maior pontuação são maiores); geralmente medido pela área sob a curva receiver-operating characteristic
Generalização	capacidade de um sistema prognóstico para fornecer previsões precisas numa nova amostra de doentes
Reprodutividade	sistemas que é preciso em doentes que não foram incluídos no desenvolvimento, mas que são de uma população idêntica

Transportabilidade	<p>sistema que é preciso em doentes provenientes de uma população diferente, mas relacionada, ou em dados colhidos usando métodos que diferem dos utilizados no desenvolvimento;</p> <p><b>a. Histórica:</b> precisão mantida quando o sistema testou dados em tempos diferentes;</p> <p><b>b. Geográfica:</b> precisão mantida quando o modelo é testado em dados de diferentes locais</p> <p><b>c. Metodológica:</b> precisão mantida quando o sistema é testado em dados colhidos usando diferentes métodos;</p> <p><b>d. Espectro:</b> precisão mantida em amostra de doente que, em média, é mais ou menos avançada no processo da doença ou que possui um processo ou trajetória de doença um tanto ou quanto diferente;</p> <p><b>e. Intervalo de follow-up:</b> precisão mantida quando o sistema é testado durante um período mais longo ou mais curto</p>
--------------------	---

Na avaliação do desempenho dos modelos (Covinsky e Berlim 1999) devem ser considerados (1), a precisão (se as previsões coincidem com os resultados na amostra original de doente), avaliado por exame de calibração e discriminação; e a generalização (em que medida o modelo fornece previsões precisas em diferentes amostras de doentes).

Um sistema que não discrimina em estudos subsequentes provavelmente não estará calibrado (Harrell et al, 1996; Ash e Schwartz, 1994), mas muitos modelos perdem a calibração enquanto mantêm alguma discriminação. Como a importância relativa da calibração e discriminação dependem do pretendido, é importante testar ambos. Além disso, a generalização de um sistema preditivo depende da sua reprodutibilidade e transportabilidade. A reprodutibilidade pode ser avaliada usando técnicas de bootstrapping na amostra de desenvolvimento, mas a avaliação da transportabilidade exige novos dados.

Quanto maior forem os dados usados no desenvolvimento do modelo, maior é a sua capacidade para identificar as dimensões de risco importantes (variáveis preditores) e quantificar com precisão o risco

A generalização, também conhecida como validade externa (Ash e Schwartz, 1994; Knaus et al, 1985), é a capacidade de um sistema prognóstico em oferecer previsões precisas em amostras posteriores de doentes.

A validação externa do modelo de prognóstico é dificultada quando a sua utilização é diferente da que foi projetado. Por exemplo, Charlson e colaboradores (1987) desenvolveram

o índice de comorbidade para prever a mortalidade de 1 ano, usando informações num coorte de 604 doentes internados no serviço de medicina de um Hospital da Nova York durante 1 mês, em 1984, e na sua validação usaram informações de 685 doentes tratados entre 1962 e 1969 por cancro da mama no Hospital Yale New Haven, o que é evidentemente muito tempo tendo em consideração a evolução dos procedimentos terapêuticos da neoplasia da mama e mudanças na população de doentes.

As técnicas de regressão logística multivariada, geralmente são utilizadas para se estabelecerem previsões sobre o resultado em análise (Iezzoni, 1997b), visto que permitem traduzir a relação existente entre uma variável dependente ou resultado e várias variáveis explicativas. Representam um método comum para a análise de relações em que a variável dependente não é contínua (Hosmer e Lemeshow, 1989). Esta é a situação mais frequente quando se pretende analisar a mortalidade hospitalar, em que a variável dependente é dicotómica, assumindo os valores 0 (alta vivo) e 1 (alta falecido) em função de um conjunto de variáveis independentes, frequentemente designadas por covariáveis (Hosmer e Lemeshow, 1989; Ruttiman, 1994).

#### **1.4. Mortalidade**

A mortalidade é utilizada como indicador de efetividade dos cuidados prestados e indesejado para o doente, prestadores de cuidados de saúde, gestores e a sociedade em geral.

A utilização das taxas de mortalidade ajustadas pelo risco poderá revelar de grande utilidade na avaliação da efetividade dos cuidados prestados, desde que a gravidade do doente seja devidamente avaliada na admissão (Jencks et al., 1988).

Outro aspeto complexo na definição de resultados em saúde respeita à janela de observação. (Iezzoni, 1995). Não existe na literatura consenso no que respeita ao intervalo de tempo a considerar. São observadas para tal a mortalidade dentro do episódio de internamento, a 7 dias, 30 dias, 6 meses ou mais, após a admissão (Drye et al., 2012; OCDE, 2013; Park et al., 2013; Häkkinen et al., 2014). Apesar das vantagens de uma ou de outra abordagem, existem evidências que a diferença entre os resultados das medidas a 30 dias ou dentro do episódio hospitalar não são significativas (Häkkinen et al., 2015).

Segundo Lopes (2010) a medida - mortalidade hospitalar dentro do episódio de internamento - apresenta uma série de vantagens face a outras que envolvam períodos de observação fora do hospital:

- Ser fácil de obter. No caso de outros períodos está dependente de informação após a alta;
- Incluir todos os doentes independentemente da duração de internamento;
- Captar o que efetivamente ocorreu dentro do episódio de internamento, sendo comparável com outros hospitais.

#### **1.4.1. Mortalidade nos DRG e Disease Staging**

Como referido anteriormente tanto os APR-DRG como o *Staging*, disponibilizam uma escala preditiva para o risco de mortalidade.

Para os APR-DRG, os níveis de mortalidade, entendidos como “*a probabilidade de ocorrer a morte do doente*”, são constituídos por quatro classes com o seguinte escalonamento (Averill *et al.*, 1997, 2013):

- Nível 1 – menor;
- Nível 2 – moderado;
- Nível 3 – Major;
- Nível 4 – Extremo

De salientar, que a escala é ordinal e como tal não permite quantificar a magnitude das diferenças entre níveis. (Averill *et al.*, 2013).

O processo de atribuição de um nível de mortalidade a um episódio de internamento e por sua vez a um determinado GDH abrange três fases: Fase I, determinação do nível de cada diagnóstico secundário; Fase II: determinação duma subclasse de base do doente com base em todos os diagnósticos secundários; Fase III, a subclasse final do doente é determinada pela incorporação do impacto do diagnóstico principal, idade, procedimento

realizado no bloco operatório, procedimento não realizado no bloco operatório de cirurgia, procedimentos não relacionados, procedimentos múltiplos realizados no bloco operatório e combinações de categorias de diagnósticos secundários. (Averill *et al.*, 1997).

Por sua vez, o *Staging* além de atribuir uma escala ordinal de gravidade à doença principal e às comorbilidades, atribui também, a cada episódio, numa escala de razão, um índice global de gravidade. (Medstat, 2010).

A escala de razão contrariamente as escalas ordinais, permite comparar entre si, episódios com diferentes doenças e comorbilidades. Possibilita também, desde que se obtenha a discriminação do modelo, conhecer para cada episódio a probabilidade de morte. (Iezzoni, 2003).

Para a atribuição da referida escala o *Staging* recorre a modelos de regressão logística. Para operacionalização dá-se os seguintes passos: separação os episódios de internamento em cirúrgicos e dos não cirúrgicos. A seguir as doenças onde for apurado menos de 300 admissões, o valor esperado corresponde ao valor observado. Para as doenças com mais de 300 admissões as regressões logísticas usam como variáveis independentes: a doença principal e respetivo estágio, a idade, o sexo, as comorbilidades e o tipo de admissão (urgente ou não urgente). O procedimento é semelhante tanto para os casos cirúrgicos como para os não cirúrgicos, a única diferença é que para os casos cirúrgicos o *Staging* usa para os cálculos tanto a doença principal como o próprio GDH. (Medstat, 2010).

Para além das dificuldades patenteadas pelas duas escalas do APR-DRG e do *Staging*, quando se pretende medir a mesma realidade, exhibe outras diferenças além da escala que é a maneira de medir. Neste curso há a destacar o efeito dos diagnósticos secundários que podem ser catalogados como comorbilidades, complicações ou ser simplesmente sintomas.

Entende-se por comorbilidades todos os diagnósticos secundários que não estão relacionados com o diagnóstico principal (Iezzoni, 1997b). Geralmente estão presentes na admissão no hospital e são frequentemente considerados indicadores válidos na previsibilidade de morte a médio e longo prazos (Charlson *et al.*, 1987) e a curto prazo (Knaus *et al.*, 1981 e 1985; Le Gall, Lemeshow e Saulnier, 1993; Lemeshow *et al.*, 1993).

Nos APR-DRG, uma complicação substancial ou comorbilidade é definida como uma condição, que devido à sua presença com um diagnóstico principal específico, obriga a um aumento do internamento, em pelo menos um dia, em 75% dos doentes. (Averill et al., 2013).

Na prática, significa que os APR-DRG não fazem a distinção entre comorbilidades e complicações, isto é, consideram os diagnósticos secundários como um todo, de acordo com o que está determinado. Este facto tem repercussão no nível do risco de morte atribuído, uma vez que este só deve considerar os elementos de admissão e não valorizar as ocorrências verificadas durante o internamento. Ao acolher as intercorrências ocorridas durante o internamento está a presentear o risco intrínseco à hospitalização, em vez de considerar somente o risco intrínseco a cada doente (procura). (Costa, 2005; Costa, Lopes e Santana, 2008).

Em oposição aos APR-DRG o Staging, faz a destrição entre diagnósticos secundários e a doença principal, possibilitando assim determinar quais são as comorbilidades e as complicações. (Medstat, 2010).

#### **1.4.2. Mortalidade no APACHE II e SAPS II**

Os modelos apresentam os resultados sob a forma de score de gravidade, traduzido por um número (geralmente mais elevada nas condições mais graves) e uma probabilidade calculada de mortalidade hospitalar.

As variáveis utilizadas são agrupadas em cinco dimensões: idade, comorbidades, alterações fisiológicas agudas, diagnóstico agudo (motivo de admissão) e intervenções (Bouch e Thompson, 2008). O diagnóstico agudo é necessário no APACHE II para obtenção da probabilidade de morte aplicando uma equação de regressão logística, mas no SAPS II é dispensada.

Quer no APACHE II quer no SAPS II, o cálculo do score de gravidade de cada doente é avaliado, utilizando um programa de computador devidamente testado e validado na sequência da atribuição a cada variável uma determinada pontuação, dependente do seu grau de afastamento dos valores normais. A soma destas pontuações constitui o valor do score de gravidade para cada doente. Estes scores estão validados para populações de doentes e não para doente, individual. No APACHE II e no SAPS II é utilizada uma escala ordinal

## 2. Objetivos

De forma genérica analisar e comparar quatro modelos em relação gravidade e probabilidade de morte.

As medidas de resultados são organizadas em três grandes grupos: medidas de efetividade (mortalidade e morbidade), medidas de eficiência (demora média e custo médio) (Iezzoni, 1997a) e medidas de adequação (adequação do procedimento e readmissões).

Na maioria dos estudos de mortalidade hospitalar, são tidas em consideração as diferenças clínicas, demográficas e de *casemix* dos doentes à admissão hospital, antes de avaliar a magnitude de quaisquer diferenças inter-hospitalar (Thomas et al, 1998) porque podem influenciar o risco de morrer no hospital (Ben-Tovim et al, 2009).

O presente estudo tem como objetivo geral avaliar quatro sistemas de classificação de doentes, na medição da gravidade da doença e da probabilidade de mortalidade hospitalar.

Assim a presente investigação tem como objetivos específicos:

- Comparar a mortalidade observada com os níveis de mortalidade preconizados pelos APR-DRG e com a escala preditiva de mortalidade do Disease Staging;
- Comparar a mortalidade observada com níveis de mortalidade preconizados pelo APACHE II e SAPS II;
- Analisar as diferenças entre os quatro sistemas de classificação de doentes na atribuição de probabilidade de morte por níveis de gravidade;
- Identificar e analisar possíveis fatores responsáveis pelas diferenças na atribuição da probabilidade de morte.



### 3. Metodologia

As fontes são: (1) dados dos resumos de alta hospitalar e (2) dados dos resumos de alta das UCI de episódios de internamento de três anos (2013, 2014 e 2015) do HSJ – Centro Hospitalar Lisboa Central. Comparou-se os níveis de mortalidade atribuídos pelos APR-DRG versão 31 com o índice de gravidade atribuído pelo software do Disease Staging, versão 5.26 e com os índices de gravidade do APACHE II e do SAPS II. Para concretização, criaram-se níveis de gravidade semelhantes aos níveis de mortalidade. Com base em regressões logísticas binárias calcularam-se e compararam-se os odd's ratios, bem como probabilidades de morte calculadas com base nas escalas de severidade e mortalidade do APR-DRG, na escala de gravidade do Staging, e nas escalas de gravidade do APACHE II e SAPS II para os episódios eleitos. Foram ainda comparados os níveis de mortalidade e gravidade com a taxa de mortalidade observada para diversos DRGs.

Em concordância com os objetivos definidos para o presente estudo, a metodologia integra um total de cinco capítulos, definidos como:

- Introdução à metodologia do estudo;
- Fontes de dados;
- População em estudo e critérios de exclusão;
- Variáveis em estudo;
- Metodologia e Operacionalização.

#### 3.1. Introdução a metodologia do estudo

Os quatro modelos que vão ser utilizados no estudo medem a severidade da doença dos doentes admitidos nos hospitais. Os momentos das colheitas dos dados e da determinação da gravidade não são coincidentes temporalmente. No APACHE II e SAPS II os dados são prospetivos, recolhidos nas primeiras 24 de internamento dos registos clínicos (Knaus et al, 1985) ainda sem efeito das medidas terapêuticas instituídas. No caso do *Disease Staging* (Thomas et al, 1986) e APR-DRGs (Averill, et al, 1998) a recolha de dados é retrospectiva, feita a partir do resumo de alta do doente, após a alta hospitalar.

O *Disease Staging* e os APR-DRG usam dados de resumos de alta-padrão, que inclui idade, sexo do doente, diagnósticos e procedimentos codificados usando o CID-9-CM. O resumo de alta contém todos os diagnósticos tratados durante a hospitalização, independentemente do momento em que os diagnósticos foram feitos.

O APACHE II e o SAPS II usam dados clínicos extraídos dos registos médicos para calcular a gravidade da doença, através do somatório das alterações fisiológicas agudas, sexo, idade as doenças crónicas graves pré-admissão, tipo de doente e motivo de admissão. Esta última variável é imprescindível para converter o score em probabilidade de morte somente no APACHE II. Os cálculos são feitos nas primeiras 24 horas de internamento

### **3.2. Fonte de dados**

As fontes de informação utilizadas na elaboração deste trabalho resultaram da (1) base de dados dos resumos de alta hospitalares (BDRAH) onde constam dados administrativos referentes aos episódios de internamento do Centro Hospitalar Lisboa Central-EPE, cedida pelo Centro de Codificação do Centro Hospitalar Lisboa Central, (2) base de dados dos resumos de alta das UCIs (BDRAUCIs): Unidade de Urgência Médica (UUM) e Unidade de Cuidados Intensivos Polivalente (UCIP1) do Hospital de São José do Centro Hospitalar Lisboa Central-EPE e, (3) do software do *Disease Staging*, da Escola Nacional de Saúde Pública para tratamento dos dados extraídos dos resumos de alta dos hospitais embora com *outputs* diferentes.

### **3.3. População em Estudo e Critérios de Exclusão**

O estudo engloba dados de três anos, respeitantes a episódios de doentes no Centro Hospitalar Lisboa Central e que tiveram internamento nas UCIs aos quais foram ainda incorporados alguns critérios de exclusão.

A população em estudo corresponde a episódios de internamentos de doentes nos anos de 2013, 2014 e 2015 com idade igual ou superior a 18 anos.

A utilização do período temporal de três anos segue os padrões habituais para atenuar eventuais valores extremos (Iezzoni et al, 1997c; O'Muircheartaigh, Murphy e Moore, 2002)

Para facilitar a compreensão dos termos utilizados, vai ser considerados como sinónimos: “sistemas”, métodos”; “doentes”, “admissões”, “casos” “episódios de internamento”; ainda “severidade” e “gravidade”. Outras designações como DRG e APR-DRG referem-se aos produtos identificados pelos *Diagnosis Related Groups*; Doença Principal, produtos identificados pelo *Disease Staging* e Diagnóstico Principal, quando se refere a CID-9-MC, “Diagnóstico agudo” e motivo de admissão APACHE II e SAPS II

No estudo foram excluídos doentes que não foram incluídos no desenvolvimento do APACHE II e SAPS (doentes pós cirurgia cardíaca, doentes coronários, doentes com idade inferior a 18 anos e doentes queimados) e doentes com internamento inferior a 24 horas na UCI por não serem portadores de score de gravidade. Também foram excluídos episódios de doentes com dados incompletos, alta contra parecer médico, episódios com informação inválida, doentes transferidos para unidade de longa duração e manutenção, unidade de cuidados paliativos e cuidados continuados integrados. Nos reinternamentos nas UCIs no decurso do episódio de internamento só foi considerado o primeiro. No total foram excluídos 852 doentes.

### **3.4. Variáveis em Estudo**

A variável dependente do estudo é a mortalidade.

As variáveis independentes referem-se a características dos doentes. São elas:

- Demográficas: idade e género;
- Tipo de doente;
- Tipo de admissão;
- Tipo de tratamento;
- DXCATS 1: Doença principal codificada de acordo com o Disease Staging;
- DXCATS 2 a 20: Comorbilidades codificadas de acordo com o Disease Staging;
- S1: Estádio e subestadio da doença principal – Disease Staging;
- S2 a 20: Estádio e Subestadio de Comorbilidades – Disease Staging;
- Gravidade: Índice de Gravidade atribuído pelo Disease Staging;

- Níveis de Gravidade: Níveis criados com base na ordenação crescente da variável gravidade e estabelecidos de acordo com os percentis dos níveis de mortalidade calculados para cada GDH;
- Comorbilidades;
- Diagnóstico Principal, no APR-DRG 31;
- Diagnósticos Secundários, APR-DRG 31;
- Doença crónica prévia, no APACHE II e SAPS II;
- Mortalidade observada;
- Mortalidade esperada.

### 3.4.1. Definições concetuais

Seguidamente encontram-se anunciadas as definições concetuais das variáveis em estudo.

- 1- **Género:** As diferenças entre os sexos podem constituir um fator adicional a considerar na avaliação do risco de morte devido das diferenças anatómicas, fisiológicas e hormonais existentes (Iezzoni, 1997d), embora a grande maioria dos sistemas desenvolvidos não inclui o sexo como fator de risco de morte. (Iezzoni, 1997b).
- 2- **Idade:** Considerada um fator de risco independente das outras características dos doentes, na medida em que a reserva fisiológica diminui com o aumento da idade. Na grande maioria das situações clínicas, os idosos apresentam maiores probabilidades de desenvolverem situações adversas do que a população jovem (Iezzoni, 1997b). Esta posição está bem patente na grande maioria dos sistemas desenvolvidos como APACHE, MPM, SAPS, Disease Staging (Walter, Knaus e Draper, 1983; Gonnella et al, 1984).
- 3- **Tipo de admissão:** Consideradas as admissões programadas e urgentes. Costa, Lopes e Santana (2008) referem que a forma de contacto com as organizações de saúde pode justificar riscos de morte distintos. Os doentes admitidos de urgência são mais propensos a ter níveis de gravidade e taxas de mortalidade mais elevadas do que as admissões programadas (Gonnella et al., 1984).
- 4- **Tipo de doente:** Cirúrgico (urgente e programado) e médico. O cirúrgico urgente tem geralmente gravidade e risco de morte mais elevados que os doentes cirúrgicos programados e médicos (Knaus et al, 1985). Segundo Gonnella e colaboradores (1984),

os casos cirúrgicos têm permanências consideravelmente mais longas do que os casos não cirúrgicos.

- 5- **Doença crônica prévia:** Considerada na atribuição de pontuação na avaliação da gravidade da doença (Knaus et al, 1985). Para que tal seja considerada tem de apresentar compromisso funcional definidos no desenvolvimento do APACHE II e SAPS II dos seguintes órgãos: fígado, cardiovascular, respiratório, renal e estado de imunodeficiência;
- 6- **Tipo de tratamento:** Considerado o tratamento médico e o tratamento cirúrgico, mediante a tipologia dos GDH. Costa, Lopes e Santana (2008) referem que o tipo de tratamento tem sido apontado como um fator que pode justificar riscos de morte distintos. Nas escalas preditivas de mortalidade do MEDSTAT foram separadas os DRGs médicos dos cirúrgicos porque os procedimentos cirúrgicos foram considerados um importante preditor de mortalidade intra-hospitalar.
- 7- **Destino após alta:** Morto ou vivo (para o domicílio). Mediante a verificação dos critérios de exclusão, a variável destino após alta “dsp” adquire dois resultados.
- 8- **Diagnóstico Principal:** é a hipótese relacionada com natureza da doença, que suporta as decisões terapêuticas e a recuperação do doente. Portanto, os diagnósticos condicionam a tecnologia de tratamento e consequentemente o tipo e intensidade de cuidados necessários para tratar uma doença (Hornbrook 1982). Consta do resumo de alta, sendo codificado através do CID-9-MC).
- 9- **Diagnósticos Secundários:** Consistem em outros diagnósticos, além do Diagnóstico Principal tratados durante o internamento registados no resumo de alta, até ao limite de 19, definidos como comorbilidades, complicações ou outras situações clínicas relacionadas, consideradas pertinentes. Podem ser ou não complicação do Diagnóstico Principal. Quando não está relacionado com o Diagnóstico Principal constitui comorbilidade (Iezzoni, 1997d).
- 10- **GDH:** Produtos que agrupam os episódios de internamento em categorias homogêneas em relação consumo de recursos medido pelos dias de internamento (Fetter et al., 1980). O DRG atribuído a cada internamento hospitalar, relaciona-se com o *casemix* hospitalar (isto é, os tipos de doentes tratados no hospital) (Bentes, Gonçalves, Tranqueada e Urbano, 1996).
- 11- **Doença Principal:** Consiste na patologia apontada pelo *Disease Staging* como o responsável pelo internamento, com base na informação do resumo de alta “dxcats1. Há uma e só uma doença principal para cada episódio de internamento (MEDSTAT, 2010).

**12- Estádio e subestádios da doença principal:** O *Disease Staging* identifica os Estádios e Sub-estádios de gravidade, que representam extensão da progressão da doença:

- Estádio 1 – Doentes menos graves;
- Estádio 2 – Doentes com gravidade intermédia;
- Estádio 3 – Doentes mais graves.
- Estádio 4 – Morte.

**13- Gravidade/severidade dos casos:** Definido como “Probabilidade de morte ou de falência de um órgão” (Thomas, Ashcraft e Zimmerman, 1986). O *Disease Staging* permite conhecer o estágio da doença principal e de comorbilidade se existir. Contudo, os estádios não permitem ter uma indicação acerca da gravidade global do doente, por duas razões: (1) o estágio refere-se individualmente a cada doença, não considerando as restantes que possam estar presentes; (2) o estágio não é comparável entre doenças: o estágio 1 significa um menor risco de morte que o estágio 2 da mesma doença, mas não é possível compará-lo ou hierarquizá-lo com os estádios de outras doenças (Gonnella, Hornbrook e Louis, 1984). Portanto, para definir um nível de gravidade global utiliza-se a informação da probabilidade de morte individual, desde que se obtenha a discriminação do modelo (Ash e Shwartz, 1997), pode ser usada para esse efeito.

**14- APR-DRG:** Tem 2 subgrupos: severidade e risco de morte da doença. A severidade é entendida como a extensão de uma descompensação fisiológica ou da perda de funções de um órgão, com 4 níveis: menor, moderado, major, extremo (Averill et al, 1998).

**15- Acute Physiology Score:** Pontuação das alterações fisiológicas agudas do doente nas primeiras 24 horas de internamento na UCI. (Knaus, et al., 1985).

**16- Índice do APACHE II:** Gravidade da doença que é obtido do somatório de pontos atribuídos às alterações fisiológicas agudas apresentadas pelo doente nas primeiras 24 horas de internamento, à idade e as doenças crónicas graves, pré-admissão, motivo de admissão até um total de 71 pontos (Knaus et al, 1985).

**17- Índice do SAPS II:** Gravidade da doença que é obtido do somatório de pontos atribuídos às alterações fisiológicas agudas apresentadas pelo doente nas primeiras 24 horas de internamento, idade, tipo de admissão e 3 variáveis relacionadas com doença subjacente. A pontuação varia de 0 e 163 pontos. (Le Gall, Lemeshow e Saulnier, 1993).

**18- Comorbilidades:** Número de diagnósticos secundários que não estão relacionados com a doença principal e que, por isso, são considerados pelo *Staging* como co-

morbilidades (Costa, Lopes e Santana, 2008; Lopes, 2010) que podem levar a um maior consumo de recursos devido ao maior risco de morte e de ocorrência de complicações, bem como a existência de reservas fisiológicas mais baixas (Iezzoni, 1997; Valderas et al., 2009). Para efeitos deste estudo, designaram-se por comorbilidades as doenças que podem ser atribuídas pelo *Disease Staging* além da doença principal, até ao limite de 19.

19- **DSP's:** Código de destino Após a Alta.

20- **Mortalidade:** Indicador de efetividade dos cuidados prestados e indesejado para o doente, prestadores de cuidados de saúde, gestores e sociedade em geral. A mortalidade pode ser um indicador de qualidade quando está ajustada às características clínicas dos doentes (Thomas et al, 1998; Ben-Tovim et al, 2009).

21- **Mortalidade observada:** Apurada através da informação presente na base de dados dos resumos de alta hospitalares e dos registos médicos. Esta permite distinguir quais os doentes que faleceram e os outros.

22- **Mortalidade esperada:** Obtida da soma das probabilidades de morte individual de todos os doentes que integrem o mesmo grupo (GDH, doença principal).

O *Staging*, para cada episódio de internamento, disponibiliza o risco de morte individual, calculado a partir da informação referente a cada doente, a enumerar: doença principal, comorbilidades, sexo, idade, tipo de admissão e procedimentos realizados. (Thomson Reuters, 2009).

Para atribuir o referido risco de morte o *software* do *Staging* recorreu à análise de uma base de dados com aproximadamente 15 milhões de registos de doentes para os quais o resultado da alta é conhecido (alta vivo/óbito). No entanto, as previsões daí obtidas não podem ser generalizadas para outras bases de dados, espaço e temporalmente distintas. Portanto, os dados originais da amostra em estudo exigem recalibração, processo explicado adiante. (Iezzoni, 2003; Justice, Covinsky e Berlin, 1999).

### 3.4.2. Definições operacionais

**TABELA 4 DEFINIÇÕES OPERACIONAIS DAS VARIÁVEIS DO ESTUDO FONTES: BDRAH E BDRAUCIS**

Nome	Descrição
<b>GÉNERO</b>	Corresponde ao género do doente: Masculino, Feminino
<b>IDADE</b>	Idade do doente, em anos, à data da admissão no internamento
<b>FAIXA ETÁRIA</b>	Agrupamento da variável idade nas seguintes classes:
	18-24 anos; 25-34 anos; 35-44 anos; 45-54 anos; 55-64 anos;
	65-74 anos; 75-84 anos; ≥ 85 anos
<b>OUTCOME</b>	Corresponde ao estado do doente aquando da alta: 0 - Alta Vivo; 1 - Óbito
<b>TIPO DE ADMISSÃO</b>	Modo de admissão do doente no hospital: 1 - Programada; 2 - Urgente
<b>GCD</b>	Código de Grande Categoria de Diagnóstico onde o episódio de internamento foi agrupado de acordo com a versão 31 dos APR-DRG. As categorias são exclusivas entre si e correspondem a sistemas orgânicos ou etiologias.
<b>GDH</b>	Código de Grupo de Diagnóstico Homogéneo (GDH) em que o episódio foi agrupado de acordo com a versão 31 dos APR-DRG.
	Tipo de GDH: M ou C
<b>TIPO DE GDH</b>	M - Médico (quando não existiram intervenções cirúrgicas) C – Cirúrgico (quando existiram intervenções cirúrgicas)
<b>NIVEIS DE MORTALIDADE</b>	Nível de risco de mortalidade atribuído a cada episódio de acordo com a versão 31 dos APR-DRG:
	1 - Menor; 2 - Moderado; 3 - Major; 4 - Extremo
<b>CÓDIGO DE DIAGNÓSTICO PRINCIPAL</b>	Código da ICD-9-CM ("International Classification of Diseases, 9th Revision, Clinical Modification"), que identifica o diagnóstico principal.
<b>CÓDIGOS DE DIAGNÓSTICOS SECUNDÁRIOS</b>	Códigos da ICD-9-CM ("International Classification of Diseases, 9th Revision, Clinical Modification"), que identifica um ou mais diagnósticos adicionais atribuídos a um doente, num determinado episódio de cuidados, para além do diagnóstico principal.
<b>CÓDIGO DE PROCEDIMENTO</b>	Código da ICD-9-CM ("International Classification of Diseases, 9th Revision, Clinical Modification") de procedimentos realizados ao doente.
<b>PRESENTE NA ADMISSÃO</b>	Identifica se o diagnóstico estava presente aquando da admissão do doente:
	S - Sim; N – Não; D - Desconhecido; I – Indeterminado NA – Isento de Registo
<b>DXCATS 1</b>	Doença principal codificada de acordo com o Disease Staging
<b>DXCATS 2 A 20</b>	Comorbilidades codificadas de acordo com o Disease Staging



	Estádio e subestádio da doença principal
<b>S2 A S20</b>	Estádio e subestadio das comorbilidades
<b>GRAVIDADE</b>	Índice de gravidade atribuído pelo Disease Staging
<b>NIVEIS DE GRAVIDADE</b>	Níveis criados com base na ordenação crescente da variável gravidade e estabelecidos de acordo com os percentis dos níveis de mortalidade calculados para cada GDH
<b>DOENÇA CRÓNICA</b>	Atribuição de pontuação na avaliação da gravidade da doença no APACHE II e SAPS II
<b>ACUTE PHYSIOLOGY SCORE</b>	Pontuação das disfunções fisiológicas agudas do doente 1 <sup>as</sup> 24 horas internamento
<b>APACHE II</b>	Pontos atribuídos as alterações fisiológicas agudas nas 1 <sup>as</sup> 24 horas de internamento
<b>APACHE II risco</b>	Conversão dos pontos em probabilidade de morte
<b>SAPS II</b>	Pontos atribuídos as alterações fisiológicas agudas nas 1 <sup>as</sup> 24 horas de internamento
<b>SAPS risco</b>	Conversão dos pontos em probabilidade de morte

### 3.5. Metodologia e Operacionalização

Com base na amostra de doentes selecionada de 1587 após aplicação dos critérios de exclusão fez-se a análise separada dos sistemas de classificação de doentes genéricos e dos sistemas clínicos de classificação de doentes.

#### ✓ Sistemas de classificação de doentes genéricos:

Análise da distribuição de cada GDH relativamente aos quatro níveis de mortalidade preconizados pelos APR- DRG e calculou-se os percentis de cada intervalo.

De seguida, dentro de cada GDH, procedeu-se a ordenação dos episódios, por ordem crescente, com base no índice de gravidade atribuído pelo Staging e criaram-se quatro níveis de gravidade observando os percentis já calculados para cada GDH.

Assim, para todos os GDH obtiveram-se quatro níveis de gravidade com base no Staging, cada um com o mesmo número de episódios que os observados pelos quatro níveis de mortalidade dos GDH.

Por último, para a obtenção dos odd's ratios, após a definição de níveis idênticos em termos numéricos, para os dois sistemas de classificação genéricos recorreu-se à regressão logística binária realizada separadamente para cada sistema. Nos APR-DRG usou-se os quatro níveis de mortalidade e no Staging os quatro níveis de gravidade anteriormente calculados.

✓ Sistemas de classificação de doentes clínicos:

A análise da distribuição de cada índice de gravidade do APACHE II relativamente aos quatro níveis de gravidade calculadas a partir dos percentis de cada intervalo.

Também no SAPS II procedeu-se a análise da distribuição de cada índice de gravidade relativamente aos quatro níveis de gravidade calculados a partir dos percentis de cada intervalo.

De seguida, avançou-se para a ordenação dos doentes, por ordem crescente, com base nos níveis de gravidade obtidos tanto no APACHE II como no SAPS II.

Logo, para todos os doentes obtiveram-se quatro níveis de gravidade com base no APACHE II, cada um com o mesmo número de doentes que os observados pelos quatro níveis de gravidade do SAPS II.

Por último, para a obtenção dos odd's ratios, após a definição de níveis idênticos em termos numéricos, para os dois sistemas de classificação de doentes específicos, recorreu-se à regressão logística realizada separadamente para cada sistema. Em ambos usou-se os quatro níveis de gravidade.

Em modo de resumo, para operacionalização a regressão cumpriu-se as seguintes formalidades:

**Variável dependente:**

- Outcome (0 em doentes com alta vivo, 1 em doentes falecidos),

**Variável independente:**

- Quatro níveis de mortalidade preconizado pelos APR-DRG;
- Quatro níveis de gravidade de acordo com o Staging;
- Quatro níveis de gravidade de acordo com o APACHE II;
- Quatro níveis de gravidade de acordo com o SAPS II.

Então, dado que as variáveis independentes eram categóricas, obtiveram-se os odd's ratios para cada nível de mortalidade e gravidade tendo como referência o primeiro nível de cada um. Em síntese, um total de odd's ratio para cada sistema de classificação.

Para a recalibração do modelo recorreu-se à regressão logística. Portanto, para se obterem as “novas” probabilidades de morte preconizadas pelo Staging, a regressão logística assentou-se nos seguintes pressupostos:

**Variável dependente:**

- Outcome (0 em doentes com alta vivo, 1 em doentes falecidos),

**Variável independente:**

- Transformação da probabilidade esperada não recalibrada no seu logit, através da fórmula:  $\log [p/1-p]$ , que corresponde à probabilidade de o evento ocorrer (p) e a probabilidade de este não ocorrer (1-p) (Costa, 2005; Lopes 2010).

No que respeita aos APR-DRG, estes disponibilizam apenas quatro níveis de severidade e quatro níveis de mortalidade, independente, para cada GDH. Como tal, para se obterem as probabilidades de morte por episódio, procedeu-se a um processo de todo idêntico, mas com pressupostos diferentes:

**Variável dependente:**

- Outcome (0 em doentes com alta vivo, 1 em doentes falecidos);

**Variável independente:**

- Transformação da probabilidade esperada não recalibrada no seu logit, através da fórmula:  $\log: [p/1-p]$ , que corresponde à probabilidade de o evento ocorrer (p) e a probabilidade de este não ocorrer (1-p)
- Outras variáveis respeitantes aos episódios: Quatro níveis de severidade, quatro níveis de mortalidade e o respetivo GDH preconizado pelos APR-DRG.

## 4. Resultados

Por forma a tornar mais clara e a facilitar a perceção deste capítulo, optou-se por dividir o mesmo em subcapítulos:

- Caracterização dos episódios de internamento;
- Comparação entre APR-DRG e *Staging*;
- Comparação entre APACHE II e SAPS II;
- Análise por níveis de mortalidade;
- Análise por níveis de gravidade.

### 4.1. Caraterização dos internamentos

Em relação à população acessível, a amostra corresponde a 1587 dos episódios de internamento

Quanto ao género, a população é maioritariamente masculina (59,00%).

Relativamente à distribuição das idades, a média da população situa-se nos 63,77% anos, mediana 67,00, anos, idade mínima 18 anos e idade máxima 95 anos.

No que respeita ao tipo de admissão, 82,0% dos episódios são urgentes e 18,0% programados (quadro anexo).

### APACHE II

**TABELAS 5 E 6 APACHE II: ESTATÍSTICAS DESCRITIVAS POR NÍVEL DE GRAVIDADE PARA VARIÁVEIS COM INTERESSE CLÍNICO**

Níveis de Gravidade	Casos	Idade		Dias de Internamento	
		Média	Desvio Padrão	Média	Desvio Padrão
1	425	55,1+0	18,22	28,20	38,43
2	374	63,89	15,97	32,34	35,97
3	404	67,68	15,28	32,76	39,90
4	384	69,14	13,63	24,44	30,27

TABELA 6

Níveis de Gravidade	Casos	Idade		Dias de Internamento	
		Mediana	AIQ	Mediana	AIQ
1	425	56	29	16	22
2	374	66	23	22	26,25
3	404	71	21,75	21	27,25
4	384	71,50	18	17	25

Para avaliar a significância das variáveis independentes, tipo de doente, sexo, idade, motivo de admissão, APACHE II, APACHE II Risco, sobre a probabilidade de ocorrer um óbito recorreu-se à regressão logística pelo método *Forward Condicional* que construiu o modelo em 4 passos. Procedeu-se também à validação dos pressupostos por intermédio do diagnóstico de casos influentes e da análise gráfica dos resíduos. Foram identificadas 31 observações que eram candidatas a *outliers*. Estas foram incorporadas no modelo final, dado que a sua remoção não melhorava nem a qualidade de ajustamento do ajustamento do modelo logístico nem a sua significância. Todas as análises foram efetuadas com o software SPSS Statistics (v.25, IMB SPSS, Chicago, IL)

A *tabela dos outliers* em anexo permite identificar as observações que podem constituir obstáculos à calibração do modelo. Todos os pontos visíveis acima da linha 4 constituem episódios que podem enviesar os resultados. Além deste aspeto, através do gráfico é possível visualizar a dimensão de cada um desses pontos que representa a estatística de influência de Cook's cuja dimensão é proporcional à capacidade de influência dessa mesma observação. A leitura da *tabela 1* em anexo como complemento do diagrama de dispersão permite visualizar os 31 *outliers* e identificar a que episódios correspondem na base de dados.

Considerando todos os preditores, o modelo de regressão logística revelou que a variável Idade ( $b_{Idade} = 0,034$ ;  $X^2_{Wald(1)} = 60,746$ ;  $p < 0,001$ ;  $OR = 1,034$ ), a variável Motivo de Admissão ( $b_{Distúrbios\ ácido\ base\ iónicos} = -1,269$ ;  $X^2_{Wald(1)} = 7,273$ ;  $p = 0,007$ ), a variável Motivo de Admissão ( $b_{Neurológico} = 1,102$ ;  $X^2_{Wald(1)} = 5,961$ ;  $p = 0,015$ ;  $OR = 3,009$ ), a variável Motivo de Admissão ( $b_{Paragem\ Cardíaca} = 1,027$ ;  $X^2_{Wald(1)} = 4,588$ ;  $p = 0,032$ ;  $OR = 2,792$ ), a

variável APACHEII ( $b_{APACHE II} = 0,058$ ;  $X^2_{Wald(1)} = 17,056$ ;  $p < 0,001$ ;  $OR = 1,060$ ), e a variável APACHEII Risco ( $b_{APACHE II Risco} = 0,016$ ;  $X^2_{Wald(1)} = 11,714$ ;  $p = 0,001$ ;  $OR = 1,016$ ) apresentam um efeito estatisticamente significativo sobre o *Logit* da probabilidade de acontecer um óbito.

Relativamente ao rácio das chances, em relação à variável idade as chances de ocorrer um óbito, aumentam 3,4% quando a variável independente varia uma unidade, a variável APACHE II as chances aumentam 6% e a variável APACHE II Risco as chances aumentam 1,6%.

Quanto à variável motivo de admissão, para a classe '*distúrbios ácido base iónicos*' as chances diminuem 71,9% quando se passa da classe de referência para a classe em teste; para a classe '*neurológico*' aumentam 200,9% e para a classe '*paragem cardíaca*' aumentam 179,2%.

Os coeficientes *Logit* do modelo de regressão logística da variável dependente em função das variáveis independentes podem ser visualizados na *tabela 2* que resume os coeficientes do modelo e a sua significância.

Assim, recorrendo ao método *Forward Conditional* a 4 passos, ajustou-se um novo modelo, estatisticamente significativo. O modelo *Logit* ajustado, apresentou os seguintes resultados ( $G^2(17) = 528,892$ ;  $p < 0,001$ ;  $X^2_{Wald}(8) = 4,280$ ;  $p = 0,831$ ;  $R^2_{CS} = 0,283$  e  $R^2_N = 0,384$  )

A percentagem de classificações correta é de 75,5%. Este valor é consideravelmente superior à percentagem proporcional de classificações pela linha de referência dos 50% e superior à percentagem proporcional de classificações corretas dadas pelo *bloco 0* do modelo (61,1%) demonstrando assim a utilidade do modelo para classificar novas observações. O modelo ajustado apresenta um valor de sensibilidade (63,4%) e elevada especificidade (83,2%) bem como uma boa capacidade discriminante ( $AUC = 0,822$ ;  $p < 0,001$ ;  $IC_{95\%} = ] 0,802; 0,842 [$ ).

**TABELA 7 APACHE II: ESTATÍSTICAS DISTRIBUÍDAS POR NÍVEIS DE GRAVIDADE, ME MORTALIDADE ESPERADA, MO MORTALIDADE OBSERVADA**

Níveis de Gravidade	Casos	Peso Relativo dos Casos	ME	MO	Razão entre MO e ME	Taxa de Mortalidade	Peso Relativo dos Óbitos
1	425	27%	60	49	0,82	12%	8%
2	374	24%	103	112	1,08	30%	18%
3	404	25%	182	190	1,05	47%	31%

4	384	24%	274	267	0,98	70%	43%
---	-----	-----	-----	-----	------	-----	-----

Na tabela 7 é visível que o modelo apresenta boas previsões para a mortalidade esperada MO versus a mortalidade observada ME, sendo que para o nível 1 de gravidade é onde os resultados dessa previsão são menos satisfatórios.

### SAPS II com introdução do variável motivo de admissão

**TABELAS 8 E 9 SAPS II: ESTATÍSTICAS DESCRITIVAS POR NÍVEL DE GRAVIDADE PARA VARIÁVEIS COM INTERESSE CLÍNICO**

Níveis de Gravidade	Casos	Idade		Dias de Internamento	
		Média	Desvio Padrão	Média	Desvio Padrão
1	420	51,88	17,04	27,95	39,25
2	391	65,24	15,45	31,61	33,06
3	391	69,15	14,24	34,28	41,48
4	385	69,78	13,61	23,90	30,31

**TABELA 9**

Níveis de Gravidade	Casos	Idade		Dias de Internamento	
		Mediana	AIQ	Mediana	AIQ
1	420	53	26,75	16	13,75
2	391	67	21	21	25
3	391	72	21	24	28
4	385	73	17,50	16	25,50



Segundo a literatura (Metnitz et al, 1999; Poncet et al 2017) a variável motivo de admissão pode ser significativa para o modelo de ajustamento pelo risco. Desta forma foi construído um novo modelo no sentido avaliar a significância das variáveis independentes, tipo de doente, sexo, idade, SAPS II, SAPS II Risco com a inclusão da variável, motivo de admissão sobre a probabilidade de ocorrer um óbito recorrendo-se inicialmente à regressão logística pelo método *Enter* e posteriormente ao método *Forward Condicional*. À semelhança dos sistemas anteriores, neste caso o método *Forward Condicional* também permitiu atingir melhores resultados. O modelo foi construído em 4 passos. Procedeu-se também à validação dos pressupostos por intermédio do diagnóstico de casos influentes e da análise gráfica dos resíduos. Foram identificadas 30 observações que eram candidatas a *outliers*. Estas foram incorporadas no modelo final, dado que a sua remoção não melhorava nem a qualidade de ajustamento do ajustamento do modelo logístico nem a sua significância.

A *tabela dos outliers* em anexo permite identificar as observações que podem constituir obstáculos à calibração do modelo. Todos os pontos visíveis acima da linha 4 constituem episódios que podem enviesar os resultados. Além deste aspeto, através do gráfico é possível visualizar a dimensão de cada um desses pontos que representa a estatística de influência de Cook's cuja dimensão é proporcional à capacidade de influência dessa mesma observação. A leitura da *tabela 5* em anexo como complemento do diagrama de dispersão permite visualizar os 30 *outliers* e identificar a que episódios correspondem na base de dados.

Considerando todos os preditores, o modelo de regressão logística revelou que a variável Idade ( $b_{Idade} = 0,028$ ;  $X^2_{Wald(1)} = 40,928$ ;  $p < 0,001$ ;  $OR = 1,029$ ), a variável SAPSII ( $b_{SAPS II} = 0,035$ ;  $X^2_{Wald(1)} = 7,486$ ;  $p = 0,006$ ;  $OR = 1,036$ ), a variável SAPS II Risco ( $b_{SAPS II Risco} = 0,016$ ;  $X^2_{Wald(1)} = 4,588$ ;  $p = 0,023$ ;  $OR = 1,016$  e a variável Motivo de admissão ( $b_{Distúrbios\ ácido\ base/iónicos} = -1,262$ ;  $X^2_{Wald(1)} = 6,917$ ;  $p = 0,009$ ) apresentam um efeito estatisticamente significativo sobre o *Logit* da probabilidade de acontecer um óbito.

Relativamente ao rácio das chances, em relação à variável idade, as chances de ocorrer um óbito, aumentam 2,9% quando a variável independente varia uma unidade, a variável SAPS II as chances aumentam 3,6% e a variável SAPS II Risco as chances aumentam 1,6%.

Quanto à variável motivo de admissão, para a classe ‘*distúrbios ácido base iônicos*’ as chances diminuem 71,7% quando se passa da classe de referência para a classe em teste.

Os coeficientes *Logit* do modelo de regressão logística da variável dependente em função das variáveis independentes podem ser visualizados na *tabela 6* em anexo que resume os coeficientes do modelo e a sua significância.

Assim recorrendo ao método *Forward Condicional* a 4 passos, conseguiu-se ajustar um novo modelo, estatisticamente significativo. O modelo *Logit* ajustado, apresentou os seguintes resultados ( $G^2(17) = 570,058$ ;  $p < 0,001$ ;  $X^2_{Wald}(8) = 11,005$ ;  $p = 0,201$ ;  $R^2_{CS} = 0,302$  e  $R^2_N = 0,409$  )

A percentagem de classificações correta é de 75,7%. Este valor é consideravelmente superior à percentagem proporcional de classificações pela linha de referência dos 50% e superior à percentagem proporcional de classificações corretas dadas pelo *bloco 0* do modelo (61,1%) demonstrando assim a utilidade do modelo para classificar novas observações. O modelo ajustado apresenta um valor de sensibilidade (63,6%) e elevada especificidade (83,5%) bem como uma boa capacidade discriminante ( $AUC = 0,832$ ;  $p < 0,001$ ;  $IC_{95\%} = [0,812; 0,851]$  ).

**TABELA 10 SAPS II: ESTATÍSTICAS DISTRIBUÍDAS POR NÍVEIS DE GRAVIDADE, ME MORTALIDADE ESPERADA, MO MORTALIDADE OBSERVADA**

Níveis de Gravidade	Casos	Peso Relativo dos Casos	ME	MO	Razão entre MO e ME	Taxa de Mortalidade	Peso Relativo dos Óbitos
1	420	26%	39	34	0,88	8%	6%
2	391	25%	105	115	1,09	29%	19%
3	391	25%	186	177	0,95	45%	29%
4	385	24%	288	292	1,01	76%	47%

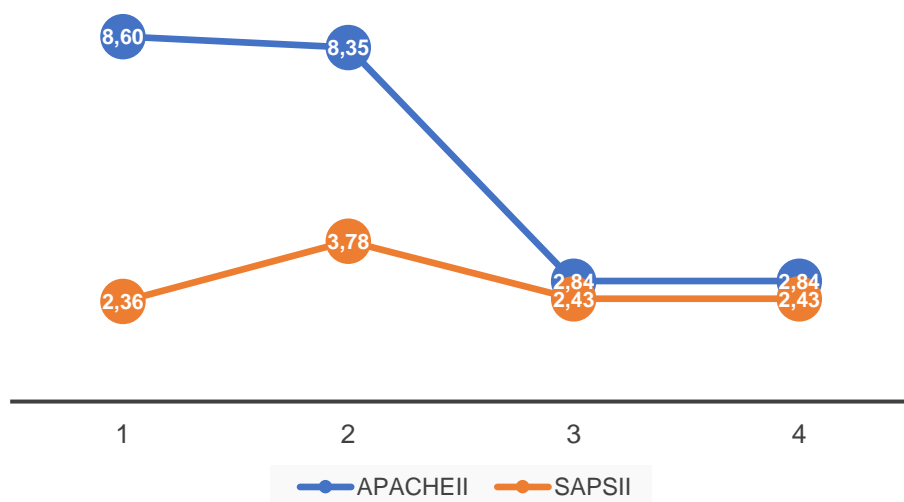
Na tabela 10 é possível concluir que o modelo apresenta boas previsões para a mortalidade esperada MO versus a mortalidade observada ME, sendo que para o nível 1 de gravidade é onde os resultados dessa previsão são ligeiramente menos satisfatórios.

## Comparação entre sistemas APACHE II e SAPS II

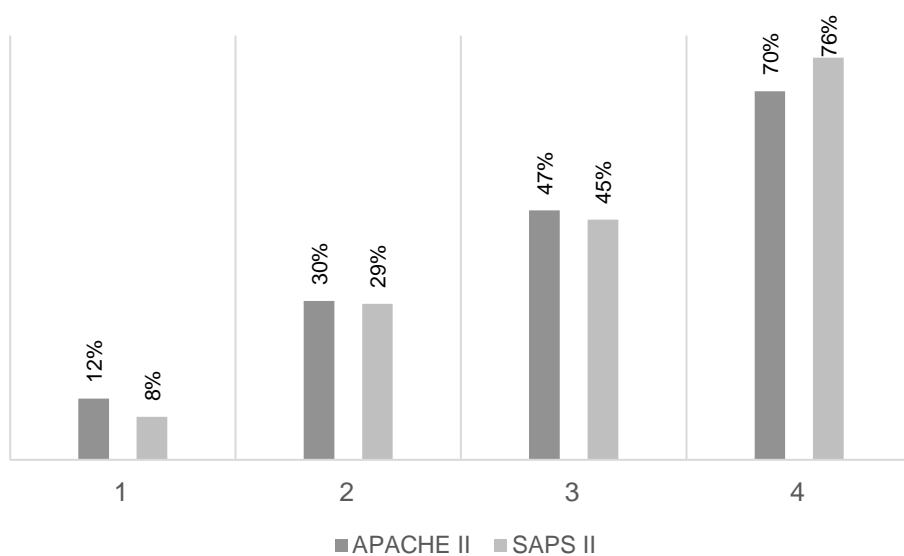


**GRÁFICO 1 COMPARAÇÃO ENTRE SISTEMAS:**  
**APACHE II vs. SAPS II ATRAVÉS DA RAZÃO ENTRE MO E ME, POR NÍVEIS DE GRAVIDADE.**

Em relação a existência de maior diferença entre a mortalidade esperada e observada (baixa mortalidade) no nível 1 do APACHE II e SAPS II poderá ser atribuído a baixa sensibilidade dos dois sistemas a este nível, bons cuidados, com maior efetividade. A outra hipótese, menos provável poderá ser a valorização do risco “oferecido” pelas unidades.



**GRÁFICO 2 COMPARAÇÃO ENTRE SISTEMAS:  
APACHE II vs. SAPS II ATRAVÉS DOS ODDS RATIOS, POR NÍVEIS DE GRAVIDADE.**



**GRÁFICO 3 COMPARAÇÃO ENTRE SISTEMAS:  
APACHE II vs. SAPS II ATRAVÉS DA TAXA DE MORTALIDADE, POR NÍVEIS DE GRAVIDADE.**

## APR-DRG

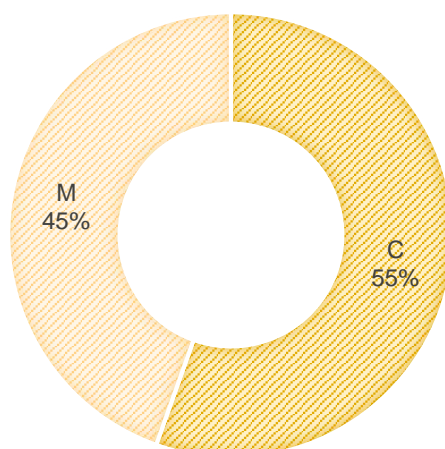


GRÁFICO 4 DISTRIBUIÇÃO DA VARIÁVEL TIPO GDH APR31 O SISTEMA APR-DRG

A variável anterior não foi usada no modelo, apenas foi estudada na perspectiva descritiva.

**TABELAS 11 E 12 APR-DRG: ESTATÍSTICAS DESCRITIVAS POR NÍVEL DE SEVERIDADE PARA VARIÁVEIS COM INTERESSE CLÍNICO**

Níveis de Severidade	Casos	Idade		N.º Comorbilidades		Dias de Internamento	
		Média	Desvio Padrão	Média	Desvio Padrão	Média	Desvio Padrão
1	75	50,37	18,61	3,87	2,00	17,47	17,34
2	212	57,55	17,98	5,19	2,50	21,51	23,05
3	461	64,66	16,83	7,03	2,71	24,78	33,87
4	839	66,05	15,44	8,16	2,70	35,06	40,83

**TABELA 12**

Níveis de Severidade	Casos	Idade		N.º Comorbilidades		Dias de Internamento	
		Mediana	AIQ	Mediana	AIQ	Mediana	AIQ
1	75	49	32	4	3	11	11

2	212	58,50	26,75	5	3	14	15,75
3	461	68	23	7	4	16	18
4	839	69	22	8	4	24	31

Para avaliar a significância das variáveis independentes, idade, gdh\_apr31, tipo de admissão, severidade\_apr31 e mortalidade\_apr31 sobre a probabilidade de ocorrer um óbito recorreu-se à regressão logística pelo método *Forward Condicional*. O modelo foi construído em 3 passos. Procedeu-se também à validação dos pressupostos por intermédio do diagnóstico de casos influentes e da análise gráfica dos resíduos. Foram identificadas 21 observações que eram candidatas a *outliers*. Estas foram incorporadas no modelo final, dado que a sua remoção não melhorava nem a qualidade de ajustamento do ajustamento do modelo logístico nem a sua significância.

A *tabela dos outliers* em anexo permite identificar as observações que podem constituir obstáculos à calibração do modelo. Todos os pontos visíveis acima da linha 4 constituem episódios que podem enviesar os resultados. Além deste aspeto, através do gráfico é possível visualizar a dimensão de cada um desses pontos que representa a estatística de influência de Cook's cuja dimensão é proporcional à capacidade de influência dessa mesma observação. A leitura da *tabela 9* em anexo como complemento do diagrama de dispersão permite visualizar os 21 *outliers* e identificar a que episódios correspondem na base de dados.

Considerando todos os preditores, o modelo de regressão logística revelou que a variável Idade ( $b_{Idade} = 0,032$ ;  $X^2_{Wald(1)} = 18,145$ ;  $p < 0,001$ ;  $OR = 1,033$ ) e a variável Mortalidade\_apr31 ( $b_{Mortalidade\_apr31} = 1,511$ ;  $X^2_{Wald(1)} = 89,115$ ;  $p < 0,001$ ;  $OR = 4,529$ ) apresentam um efeito estatisticamente significativo sobre o *Logit* da probabilidade de acontecer um óbito.

Em relação ao rácio das chances, relativamente à variável idade as chances de ocorrer um óbito, aumentam 3,3% quando a variável independente varia uma unidade. Quanto à variável mortalidade\_apr31 as chances aumentam 352,9%.

Os coeficientes *Logit* do modelo de regressão logística da variável dependente em função das variáveis independentes podem ser visualizados a partir da *tabela 10* que resume os coeficientes do modelo e a sua significância.

Assim recorrendo ao método *Forward Condicional* a 3 passos, ajustou-se um novo modelo, estatisticamente significativo. O modelo *Logit* ajustado, apresentou os

seguintes resultados ( $G^2(92) = 831,582$ ;  $p < 0,001$ ;  $X^2_{Wald}(8) = 15,7$ ;  $p = 0,160$ ;  $R^2_{CS} = 0,408$  e  $R^2_N = 0,553$  )

A percentagem de classificações correta é de 82,4%. Este valor é consideravelmente superior à percentagem proporcional de classificações pela linha de referência dos 50% e superior à percentagem proporcional de classificações corretas dadas pelo *bloco 0* do modelo (64%) demonstrando assim a utilidade do modelo para classificar novas observações. O modelo ajustado apresenta um valor de sensibilidade (77%) e elevada especificidade (85,4%) bem como uma capacidade discriminante excecional ( $AUC = 0,801$ ;  $p < 0,001$ ;  $IC_{95\%} = ] 0,781 ; 0,821 [$  ).

**TABELA 13 APR-DRG: ESTATÍSTICAS DISTRIBUÍDAS POR NÍVEIS DE SEVERIDADE, ME MORTALIDADE ESPERADA, MO MORTALIDADE OBSERVADA**

Níveis de Severidade	Casos	Peso Relativo dos Casos	ME	MO	Razão entre MO e ME	Taxa de Mortalidade	Peso Relativo dos Óbitos
<b>1</b>	75	5%	2	1	0,43	1%	0%
<b>2</b>	212	13%	21	19	0,91	9%	3%
<b>3</b>	461	29%	140	148	1,06	32%	24%
<b>4</b>	839	53%	455	450	0,99	54%	73%

O sistema APR-DRG, apresenta boas previsões para a mortalidade esperada MO versus a mortalidade observada ME, sendo que para o nível 1 de severidade é onde os resultados dessa previsão são acentuadamente menos satisfatórios. Tabela 13.

A baixa mortalidade do nível 1 de severidade do APR-DRG poderá ter a ver com o número de casos que é reduzido ou baixa sensibilidade do sistema a este nível.

O sistema APR-DRG, em relação aos 4 níveis de mortalidade não apresenta boas previsões para a mortalidade esperada MO versus a mortalidade observada ME comparativamente aos outros três modelos, sendo que para os 4 níveis de mortalidade os resultados dessa previsão são pouco satisfatórios. Tabela 14.

A determinação dos níveis de severidade e gravidade no APR-DRG faz-se de forma independente. O desempenho do APR-DRG severidade é melhor que o APR-DRG mortalidade como mais uma vez é documentado no estudo. A sensibilidade do APR-DRG de mortalidade no nível 4 é muito baixa.

**TABELA 14 APR-DRG: ESTATÍSTICAS DISTRIBUÍDAS POR NÍVEIS DE MORTALIDADE, ME MORTALIDADE ESPERADA, MO MORTALIDADE OBSERVADA**

Níveis de Mortalidade	Casos	Peso Relativo dos Casos	ME	MO	Razão entre MO e ME	Taxa de Mortalidade	Peso Relativo dos Óbitos
<b>1</b>	204	13%	6	2	0,32	1%	0%
<b>2</b>	188	12%	16	24	1,49	13%	4%
<b>3</b>	357	22%	65	90	1,38	25%	15%
<b>4</b>	838	53%	261	502	1,93	60%	81%

**TABELA 15 APR-DRG: ESTATÍSTICAS DISTRIBUÍDAS POR GDH'S, ME MORTALIDADE ESPERADA, MO MORTALIDADE OBSERVADA**

GDH	Casos	Peso Relativo dos Casos	ME	MO	Razão entre MO e ME	Taxa de Mortalidade	Peso Relativo dos Óbitos
<b>130</b>	153	28%	41	64	1,57	42%	26%
<b>139</b>	42	8%	11	17	1,58	40%	7%
<b>220</b>	87	16%	17	22	1,30	25%	9%
<b>221</b>	138	25%	34	61	1,80	44%	24%
<b>4</b>	39	7%	11	25	2,26	64%	10%
<b>5</b>	48	9%	17	41	2,39	85%	16%
<b>690</b>	40	7%	14	19	1,33	48%	8%

Na tabela 15 é possível verificar que a partir da variável GDH que não se conseguem estabelecer previsões satisfatórias acerca da mortalidade esperada MO versus a mortalidade observada ME, sendo um dos principais motivos que podem contribuir para esses resultados será a dimensão reduzida da amostra.

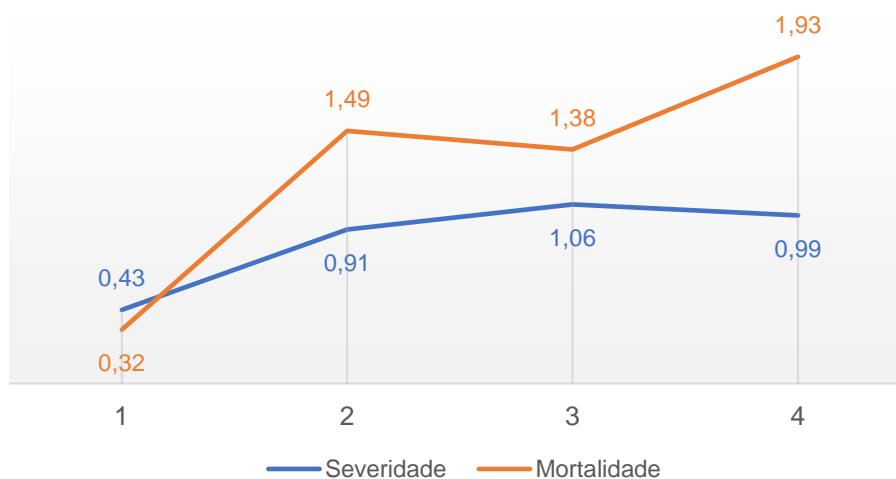


**TABELA 16 APR-DRG: DISTRIBUIÇÃO DOS CASOS POR NÍVEIS DE SEVERIDADE  
VS. NÍVEIS DE MORTALIDADE**

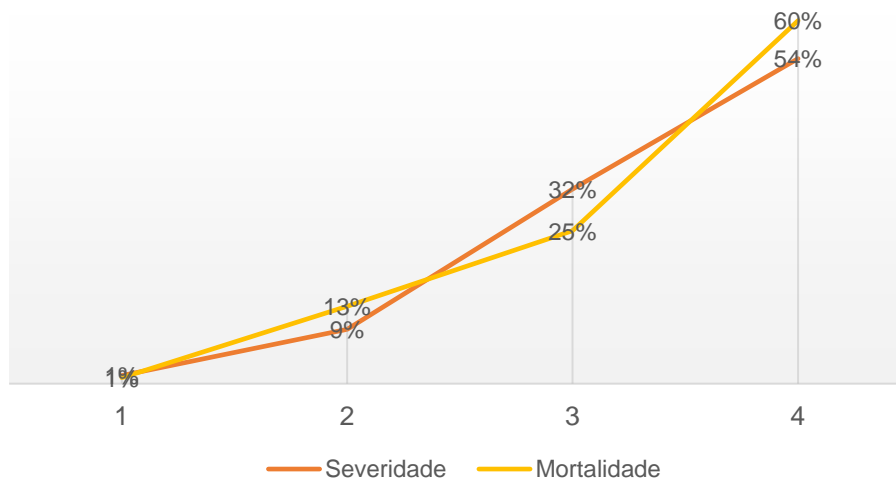
Níveis de Severidade	Níveis de Mortalidade				Total
	1	2	3	4	
1	60	14	1	0	75
2	109	63	32	8	212
3	34	99	213	115	461
4	1	12	111	715	839
<b>Total</b>	204	188	357	838	1.587

**TABELA 17 APR-DRG: PERCENTAGEM DOS CASOS POR NÍVEIS DE SEVERIDADE VS.  
NÍVEIS DE MORTALIDADE**

Níveis de Severidade	Níveis de Mortalidade			
	1	2	3	4
1	3,78%	0,88%	0,06%	0,00%
2	6,87%	3,97%	2,02%	0,50%
3	2,14%	6,24%	13,42%	7,25%
4	0,06%	0,76%	6,99%	45,05%



**GRÁFICO 5 COMPARAÇÃO ENTRE AS ESCALAS DE SEVERIDADE E MORTALIDADE ATRAVÉS DA RAZÃO ENTRE MO E ME PARA O APR-DRG**



**GRÁFICO 6 COMPARAÇÃO ENTRE AS ESCALAS DE SEVERIDADE E MORTALIDADE ATRAVÉS DA TAXA DE MORTALIDADE PARA O APR-DRG**

**Disease Staging**

**TABELAS 18 E 19 DISEASE STAGING: ESTATÍSTICAS DESCRITIVAS POR NÍVEL DE GRAVIDADE PARA VARIÁVEIS COM INTERESSE CLÍNICO**

Níveis de Gravidade	Casos	Idade		Dias de Internamento	
		Média	Desvio Padrão	Média	Desvio Padrão
<b>1</b>	396	51,04	16,80	27,46	32,89
<b>2</b>	397	61,97	16,96	32,78	40,43
<b>3</b>	398	68,61	12,59	31,62	38,15
<b>4</b>	396	73,44	11,27	25,83	33,89

**TABELA 19**

Níveis de Gravidade	Casos	Idade		Dias de Internamento	
		Mediana	AIQ	Mediana	AIQ
<b>1</b>	396	51	12	16	13
<b>2</b>	397	65	24	21	26,50
<b>3</b>	398	70,50	18	22	25,25
<b>4</b>	396	75	15	18	24

Para avaliar a significância das variáveis independentes, idade, sexo, tipo de admissão, tipo de tratamento, *dxcats1* e *logit* da estimativa de morte sobre a probabilidade de ocorrer um óbito *outcome* recorreu-se à regressão logística pelo método *Forward Condicional*. O modelo foi construído em 3 passos.

A estatística para o teste de Hosmer-Lemeshow foi igual a 5,987 com uma significância de 0,649 concluindo-se desta forma que o modelo encontra-se bem calibrado.

O modelo apresenta uma boa capacidade discriminante ( $AUC = 0,859$ ;  $p < 0,001$ ;  $IC_{95\%} = ] 0,841; 0,876 [$ ).

**TABELA 20 DISEASE STAGING: ESTATÍSTICAS DISTRIBUÍDAS POR NÍVEIS DE GRAVIDADE,  
ME MORTALIDADE ESPERADA, MO MORTALIDADE OBSERVADA**

Níveis de Gravidade	Casos	Peso Relativo dos Casos	ME	MO	Razão entre MO e ME	Taxa de Mortalidade	Peso Relativo dos Óbitos
<b>1</b>	396	25%	60	8	0,13	2%	1%
<b>2</b>	397	25%	103	86	0,83	22%	14%
<b>3</b>	398	25%	182	214	1,18	54%	35%
<b>4</b>	396	25%	274	310	1,13	78%	50%

O Disease Staging, apresenta boas previsões o nível 4 de gravidade para a mortalidade esperada MO versus a mortalidade observada ME, sendo que para os restantes níveis vai decrescendo essa probabilidade, como se pode observar através da leitura da tabela 20.

**TABELA 20 DISEASE STAGING: ESTATÍSTICAS DISTRIBUÍDAS POR BSY,  
ME MORTALIDADE ESPERADA, MO MORTALIDADE OBSERVADA**

Bsy	Casos	Peso Relativo dos Casos	ME	MO	Razão entre MO e ME	Taxa de Mortalidade	Peso Relativo dos Óbitos
CVS	157	11%	48	83	1,72	53%	14%
GIS	294	21%	66	116	1,76	39%	20%
GUS	64	5%	11	17	1,57	27%	3%
HEM	100	7%	29	51	1,78	51%	9%
HEP	72	5%	13	29	2,19	40%	5%
MUS	83	6%	6	19	2,92	23%	3%
NEU	188	13%	59	118	2,01	63%	20%
OTH	61	4%	4	4	0,95	7%	1%
RES	293	21%	75	129	1,73	44%	22%
SKN	52	4%	5	7	1,48	13%	1%
TRA	45	3%	5	8	1,73	18%	1%

Na tabela 20 é possível verificar que a partir da variável BSY que não se conseguem estabelecer previsões com qualidade acerca da mortalidade esperada MO versus a mortalidade observada ME, sendo um dos principais motivos que podem contribuir para esses resultados será a dimensão reduzida da amostra.

### **Comparação entre sistemas APACHE II e SAPS II**

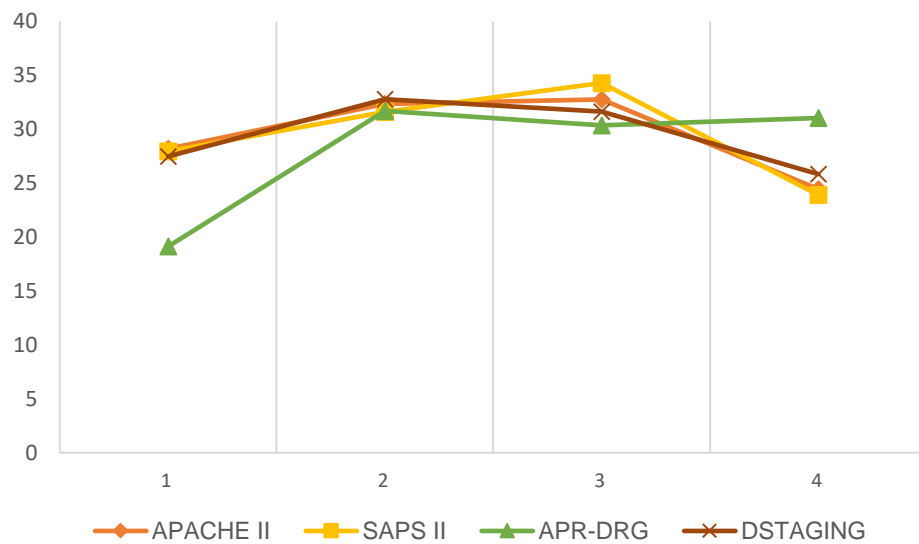
**TABELA 21 COMPARAÇÃO ENTRE NÍVEIS DE SEVERIDADE E MORTALIDADE  
DO APR-DRG E OS NÍVEIS DE GRAVIDADE DO DSTAGING**

APR-DRG						DSTAGING		
Níveis de Sev.	Razão entre MO e ME	Taxa de Mortalidade	Níveis de Mor.	Razão entre MO e ME	Taxa de Mortalidade	Níveis de Gra.	Razão entre MO e ME	Taxa de Mortalidade
<b>1</b>	0,43	1%	<b>1</b>	0,32	1%	<b>1</b>	0,13	2%
<b>2</b>	0,91	9%	<b>2</b>	1,49	13%	<b>2</b>	0,83	22%
<b>3</b>	1,06	32%	<b>3</b>	1,38	25%	<b>3</b>	1,18	54%
<b>4</b>	0,99	54%	<b>4</b>	1,93	60%	<b>4</b>	1,13	78%

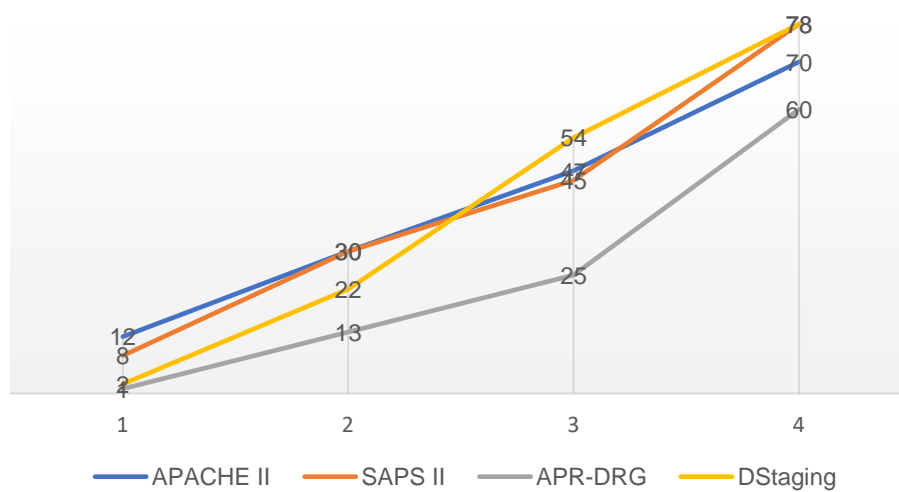
### Comparação entre os quatro sistemas APACHE II, SAPS II, APR-DRG e Disease Staging

**TABELA 23 TEMPO DE INTERNAMENTO VS PERCENTIS 4 SISTEMAS**

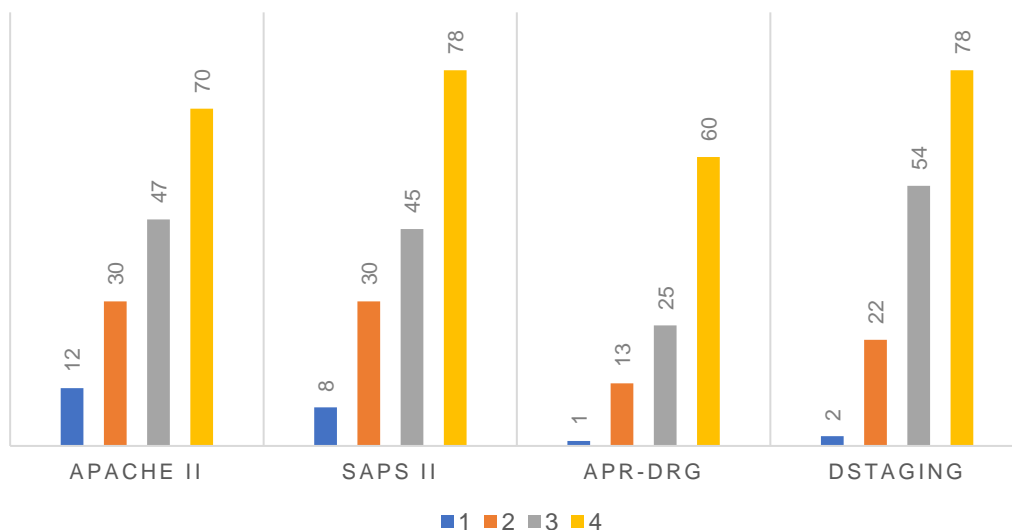
Tempo de internamento		APACHE II	SAPS II	APR-DRG	DSTAGING
<b>1</b>	Mediana ± DI	16,0 ± 22,0	16,0 ± 20,5	12,0 ± 14,0	16,0 ± 20,0
	Média ± DP	28,2 ± 38,4	27,9 ± 39,3	19,1 ± 19,5	27,5 ± 32,9
<b>2</b>	Mediana ± DI	22,0 ± 26,3	21,0 ± 25,0	19,5 ± 21,0	21,0 ± 26,5
	Média ± DP	32,3 ± 36,0	31,6 ± 33,1	31,7 ± 47,7	32,8 ± 40,4
<b>3</b>	Mediana ± DI	21,0 ± 27,8	24,0 ± 28,0	20,0 ± 22,5	22,0 ± 25,3
	Média ± DP	32,8 ± 39,9	34,3 ± 41,5	30,4 ± 39,1	31,6 ± 38,2
<b>4</b>	Mediana ± DI	17,0 ± 25,0	16,0 ± 25,5	21,0 ± 29,0	18,0 ± 24,0
	Média ± DP	24,4 ± 30,3	23,9 ± 30,3	31,0 ± 35,4	25,8 ± 33,9



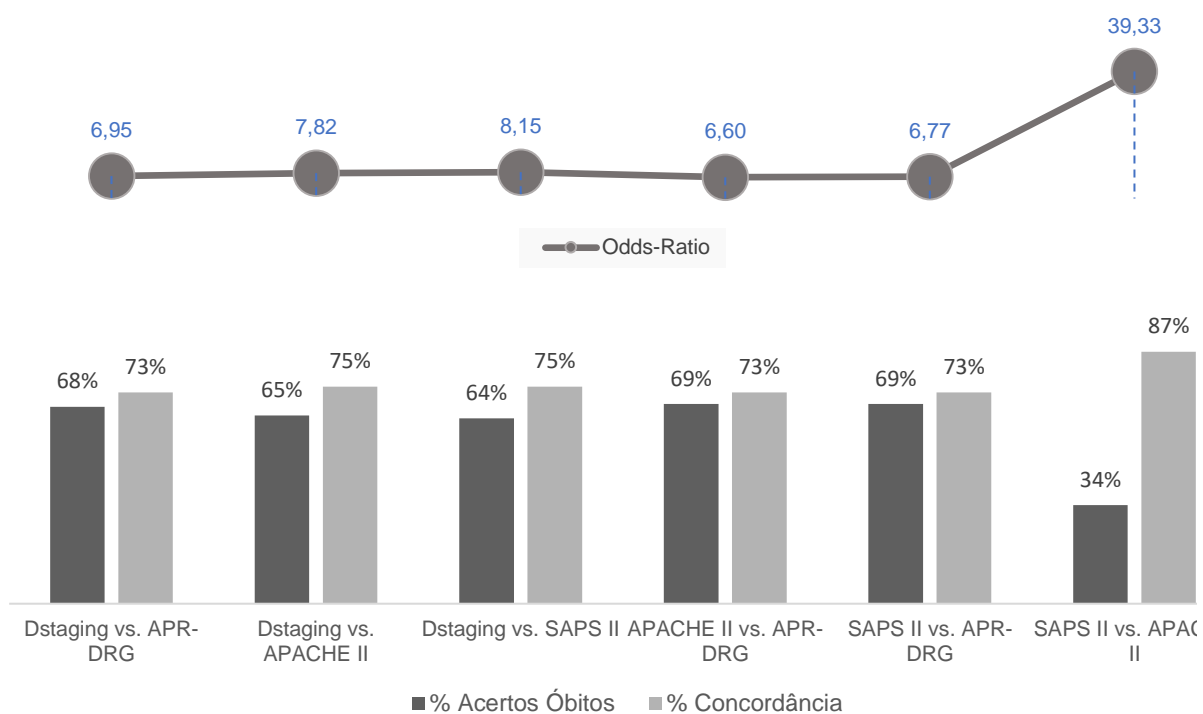
**GRÁFICO 7 MÉDIAS DO TEMPO DE INTERNAMENTO POR SISTEMA NOS 4 PERCENTIS**



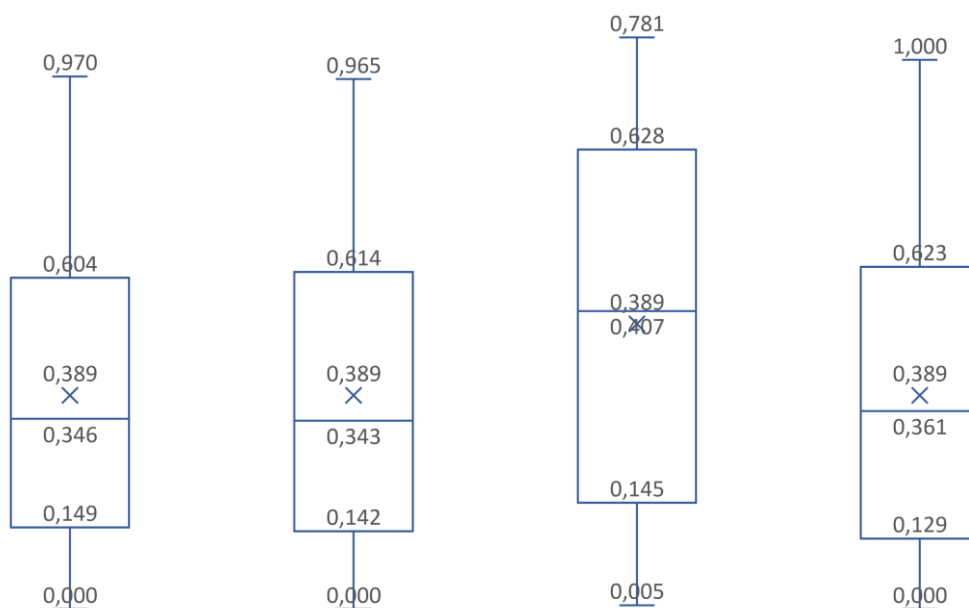
**GRÁFICO 8 PERCENTAGEM DE ÓBITOS VERIFICADOS POR PERCENTIL EM CADA SISTEMA**



**GRÁFICO 9 PERCENTAGEM DE ÓBITOS VERIFICADOS POR PERCENTIL EM CADA SISTEMA**



**GRÁFICO 10 COMPARAÇÃO DE 3 INDICADORES PARA OS 4 SISTEMAS**



**GRÁFICO 11 COMPARAÇÃO ENTRE A PROBABILIDADE DE MORTE ESPERADA PARA OS RESULTADOS DOS 4 MODELOS**

Foram também analisados os 618 óbitos, em relação aos óbitos esperados para o APACHE II, SAPS II, APR-DRG e Disease Staging, assumiram valores iguais a um, para os 618 valores mais elevados de cada um deles. Assim, em primeiro lugar contruíram-se 4 variáveis de acordo com os rankings descendentes das mortalidades esperadas para depois se poder atribuir o valor um aos 618 mais elevados e o valor zero aos restantes.

Após esta fase da análise foram realizadas tabelas de contingência entre todas as combinações possíveis, seis no total.

**TABELA 24 COMPARAÇÃO ENTRE ODDS RATIOS, PERCENTAGEM DE CONCORDÂNCIA E ACERTOS NA ESTIMATIVA DE ÓBITOS PARA OS RESULTADOS DOS 4 MODELOS.**

Sistemas		Odds-Ratio	% Concordância	% Acertos Óbitos
DStaging	APR-DRG	6,95	73%	68%
DStaging	APACHE II	7,82	75%	65%



DStaging	SAPS II	8,15	75%	64%
APACHE II	APR-DRG	6,60	73%	69%
SAPS II	APR-DRG	6,77	73%	69%
SAPS II	APACHE II	39,33	87%	34%

**TABELA 25 RESULTADOS OBTIDOS PARA OS 4 MODELOS**

	APACHE II	SAPS II	APR-DRG	DStaging
variáveis adicionadas		motivo de admissão		
Método	<i>Forward Condicional</i> 4 passos	<i>Forward Condicional</i> 4 passos	<i>Forward Condicional</i> 3 passos	<i>Forward Condicional</i> 3 passos
R <sup>2</sup> <sub>CS</sub>	0,283	0,302	0,408	na
R <sup>2</sup> <sub>N</sub>	0,384	0,409	0,553	na
Testes de Omnibus	G <sup>2</sup> (17) =528,892; <i>p</i> < 0,001	G <sup>2</sup> (17) =570,058; <i>p</i> < 0,001	G <sup>2</sup> (92) =831,582; <i>p</i> < 0,001	na
Hosmer & Lemeshow	X <sup>2</sup> <sub>Wald</sub> (8) =4,280; <i>p</i> = <b>0,831</b>	X <sup>2</sup> <sub>Wald</sub> (8) =11,005; <i>p</i> = <b>0,201</b>	X <sup>2</sup> <sub>Wald</sub> (8) =15,7; <i>p</i> = <b>0,160</b>	X <sup>2</sup> <sub>Wald</sub> (8) =5,987 <i>p</i> = <b>0,649</b>
Porcentagem Global	75,5	75,7	82,4	na
Estatística C p-value IC's 95%	<b>0,822</b> <i>p</i> < 0,001 ]0,802 ; 0,842[	<b>0,832</b> <i>p</i> < 0,001 ]0,812 ; 0,851[	<b>0,801</b> <i>p</i> < 0,001 ]0,781 ; 0,821[	<b>0,859</b> <i>p</i> < 0,001 ] 0,841 ; 0,876 [
Variáveis significativas	Idade ma: Distúrbios abi ma: Neurológico APACHE II	Idade ma: Distúrbios abi SAPS II SAPS II Risco	Idade Mortalidade_apr31	Idade

	APACHE II Risco			
Comentários	Modelo Calibrado Boa Discriminação	Modelo Calibrado Boa Discriminação	Modelo Calibrado Discriminação Excecional	Modelo Calibrado Boa Discriminação
Outliers	31	30	21	12

## 5. Discussão

Para tornar mais perceptível e facilitar a compreensão deste capítulo, obteve-se por dividir o mesmo em dois subcapítulos: a discussão metodológica e a discussão de resultados.

Para o presente estudo recorreu-se à (1) uma base de dados dos resumos de alta hospitalares (BDRAH) que é uma base de dados administrativa utilizada em Portugal para o financiamento hospitalar e (2) outra base de dados dos resumos de alta das UCIs (BDRAUCIs) que é uma base de dados clínica de duas unidades de cuidados intensivos nível III, do Centro Hospitalar Lisboa Central-EPE

O facto de os sistemas de classificação de doentes usarem bases de dados administrativa e base de dados clínicos permite comparar, e identificar as vantagens e as desvantagens.

As vantagens dos dados administrativos estão relacionadas com a facilidade de disponibilidade e acessibilidade, o que consequentemente, provoca custos operacionais mais baixos e uma menor logística em torno da colheita e tratamento dos mesmos. (Costa, 2005; DesHarnais, McMahon e Wroblewski, 1991; Iezzoni, 2003).

Em termos estatísticos os dados administrativos apresentam a mesma precisão que os dados clínicos na previsão da mortalidade hospitalar (Iezzoni et al 1995a, 1997c; 1998; Romano 1994; Pine et al 1997).

As desvantagens são várias e relacionam-se com: (1) menor precisão e validade face aos sistemas que utilizam dados clínicos (Iezzoni *et al.*, 1992, 1996); (2) a maioria

dos dados administrativos não distingue comorbidades de complicações ou entre doenças agudas e crônicas (Fokkema et al 2014); (3) não incluem critérios diagnósticos e outras informações clínicas, como pressão arterial e resultados laboratoriais (Hannan et al, 1997; Jollis et al, 1993; (4) subestimam a gravidade (Shwartz et al, 1996b); (5) são problemáticos quando utilizados na avaliação dos cuidados prestados pelo clínico, ou para comparar prestadores na perspectiva do doente ou pagador, porque podem ou não refletir com precisão os cuidados clínicos (Iezzoni, 2013); (6) dependem da documentação e das práticas de codificação das instituições, da subjetividade na escolha do diagnóstico principal e à sobreposição de categorias existente no CID-9-MC (Costa e Lopes 2004); (7) desadequados para determinar o estado dos sintomas e estados de alto risco, em grande parte devido à falta de especificidade com os códigos de diagnóstico CID-9-MC, uma vez que não conseguem fornecer informações sobre a gravidade e início temporal da doença (Bensley et al 2013); (8) ausência de informação acerca das preferências dos doentes, tal como indicação de não reanimar e pedidos apenas de medidas de conforto. Estas lacunas, podem comprometer gravemente a utilidade dos dados administrativos para produzir informações significativas sobre as taxas de mortalidade (Iezzoni LI 2012).

Segundo Krumholz e colaboradores (2006a) os dados administrativos só devem ser usados para classificar hospitais após confirmarem a sua validade para tal propósito, como aconteceu na mortalidade após EAM e insuficiência cardíaca (Krumholz et al, 2006b).

Apesar das insuficiências apontadas, os dados administrativos prestam informações importantes acerca da utilização dos serviços de saúde, despesas, seleção de resultados clínicos e qualidade dos cuidados. (Iezzoni, 2012).

Os dados clínicos além da vantagem, proporcionada pelos sinais vitais e resultados dos exames complementares de diagnóstico (Iezzoni et al., 1995b e 1996d; Iezzoni, 1997b), apresentam desvantagens que estão associadas custos, morosidade (Lawson et al 2012; Lane-Fall MB, 2013) e a perdas de informação ou a erros de transcrição (Iezzoni et al., 1995b e 1996d; Iezzoni, 1997b).

Em certas situações, por exemplo, cirurgia cardíaca os dados administrativos são combinados com dados clínicos para melhorar a mortalidade ajustada pelo risco ((Guru et al, 2008).

A codificação é um campo amplamente abordada na literatura dado a influência nos resultados finais, quer ao nível da severidade dos casos quer ao nível do risco da mortalidade. Neste âmbito são referenciados fenómeno de “*DRG creep*” (sobrecodificação) que traduz o propósito de aumentar o financiamento do hospital ou de tornar os doentes ainda mais doentes. Como corolário destes comportamentos, o risco de mortalidade é artificialmente aumentado o que leva que as probabilidades da mortalidade observadas sejam mais baixas que as esperadas, isto é o hospital apresenta melhores resultados (efetividade). (Pongpirula e Robinson, 2013).

Por outro lado, também existem estudos que asseguram que as práticas de subcodificação chegam a ser três vezes mais frequentes que o seu contrário. (Rudman et al, 2009).

A confiabilidade dos sistemas de classificação de doentes baseados em dados administrativos não foi testada diretamente, mas inferida a partir de estudos de confiabilidade de codificação (Iezzoni, 2012).

Quando chega o momento de decidir qual o sistema de classificação de doentes a utilizar no ajustamento pelo risco torna-se imperioso optar por uma das duas alternativas: “privilegiar a operacionalidade e economia dos sistemas de classificação de doentes ou privilegiar a validade de construção e de conteúdo dos sistemas de classificação de doentes?”. (Costa, 2005). No presente estudo vai ser aplicado os APR-DRG que privilegiam a operacionalidade e economia e três modelos, Disease Staging, APACHE II e SAPS que privilegiam a validade de construção e de conteúdo.

### **5.1.2. Abordagem**

#### **5.1.2.1. Sistemas genéricos de classificação de doentes**

Os dois sistemas de classificação de doentes genéricos em análise apresentam

critérios para a definição de produtos, e da escala para a probabilidade de morte, diferentes. Os APR-DRG apresentam uma escala ordinal e o *Staging* uma escala de razão.

Em correspondência com o objetivo do estudo, a única abordagem metodológica adequada e compreensível foi a de converter a escala de razão do *Staging* numa escala ordinal em tudo semelhante à dos APR-DRG, recorrendo à ordenação por ordem crescente de gravidade dos respetivos episódios.

Após ponderação concluiu-se que esta era a única que permite comparar realidades idênticas. O inverso é impossível dada as limitações inerentes a uma escala ordinal.

Contudo, a solução conseguida impõe especial cuidado na análise de resultados. O facto de se agregarem os scores de gravidade do *Staging* de acordo com os níveis de mortalidade dos APR-DRG, por forma a se obterem classes de tamanhos idênticos, pode condicionar a análise de resultados.

#### **5.1.2.2. Sistemas clínicos de classificação de doentes**

Os dois sistemas de classificação clínicos de doentes em análise apresentam os mesmos critérios para a definição de produtos, e índice de gravidade.

No APACHE II e no SAPS II é utilizada uma escala ordinal para classificar os doentes por ordem crescente de gravidade.

#### **5.1.3. Limitações operacionais**

As limitações operacionais dos quatro sistemas de classificação de doentes não são coincidentes. Os dois modelos genéricos – APR-DRG e Disease Staging foram desenvolvidos em grandes bases de dados onde em regra são utilizados. No que refere aos sistemas de classificação de doentes clínicos – APACHE II e SAPS II foram desenvolvidos em bases de dados de menor dimensão das unidades de cuidados

intensivos. Portanto, para tornar mais perceptível e para facilitar a compreensão a descrição vai ser em separado.

#### **5.1.3.1. Sistemas genéricos de classificação de doentes: APR-DRG e Disease Staging**

A escala disponibilizada pelos APR-DRG motiva limitações não só no presente estudo como quando se pretende avaliar e comparar o desempenho hospitalar.

O facto de os APR-DRG não disponibilizarem também o risco de morte individual, unicamente permite ao prestador ter conhecimento da mortalidade esperada, isto é, o prestador sabe simplesmente que um determinado episódio de um GDH em relação a outro episódio, poderá ter uma maior mortalidade esperada, em função do nível que lhe foi atribuído.

Para o cálculo da mortalidade esperada, nos APR-DRG haverá que recorrer sempre a técnicas de padronização, direta ou indireta. (Lane-Fall e Neuman, 2013).

A padronização indireta implica que a amostra de casos de um determinado hospital seja comparada com uma amostra de referência, considerada padrão. Por sua vez na padronização direta a amostra do hospital é dividida em dois ou mais subgrupos, sendo um dos grupos de validação que funciona como referência, considerada padrão.

A técnica de padronização indireta, isto é, a modelos de regressão logística é mais comum. Possibilita a obtenção da mortalidade esperada, com base em várias dimensões de risco (idade, género, diagnóstico, etc.) inferida a partir da mortalidade observada dessa mesma população.

No presente estudo os sistemas de classificação de doentes foram usados para determinar a taxa esperada de mortalidade numa população de doentes tratados no Centro Hospitalar Lisboa Central, EPE. A taxa esperada foi comparada à taxa real (observada) dos APR-DRG. Como acontece neste tipo de estudo, a comparação das

taxas observadas e esperadas de resultados foi realizada com a abordagem conhecida como padronização indireta. (Ash et al, 1997).

A operacionalização foi feita utilizando a mortalidade esperada dos dois modelos genéricos que foi posteriormente comparada com a mortalidade observada na população em estudo.

Em síntese, o cálculo da mortalidade esperada será tanto mais fiável quanto maior e mais representativa for a população que servirá de base para comparação. (Dimick et al, 2004).

#### **5.1.3.2. Sistemas clínicos de classificação de doentes: APACHE II E SAPS II**

A escala disponibilizada pelo APACHE II e SAPS II também motiva limitações não só no presente estudo como quando se pretende avaliar e comparar o desempenho hospitalar.

O facto de o APACHE II e SAPS II não disponibilizarem o risco de morte individual, unicamente permite ao prestador ter um conhecimento inexpressivo, da mortalidade esperada, isto é, o prestador sabe simplesmente que um determinado episódio de internamento em relação a outro episódio, poderá ter uma maior mortalidade esperada, em função do nível que lhe foi atribuído.

Para calcular a mortalidade esperada, no caso do APACHE II e do SAPS II haverá que recorrer sempre a técnicas de padronização, direta ou indireta. (Lane-Fall e Neuman, 2013).

A padronização indireta implica que a amostra de casos de uma unidade de cuidados intensivos de determinado hospital seja comparada com uma amostra de referência, considerada, padrão.

A técnica de padronização indireta, isto é, a modelos de regressão logística é mais comum. Possibilita a obtenção da mortalidade esperada, com base em várias dimensões de risco (idade, género, motivo de admissão, etc.) inferida a partir da

mortalidade observada dessa mesma população.

No presente estudo o APACHE II e o SAPS II foram usados para determinar a taxa esperada de mortalidade numa população de doentes tratados no Centro Hospitalar Lisboa Central, EPE. A taxa esperada de mortalidade foi comparada à taxa real (observada) dos dois modelos. Como é acontece neste tipo de estudo, a comparação das taxas observadas e esperadas de resultados foi realizada com a abordagem conhecida como padronização indireta. (Ash et al. 1997).

A operacionalização foi feita utilizando a mortalidade esperada dos dois modelos clínicos que foi posteriormente comparada com a mortalidade observada na população em estudo.

Em síntese, o cálculo da mortalidade esperada será tanto mais fiável quanto maior e mais representativa for a população que servirá de base para comparação. (Dimick et al, 2004).

O presente estudo tem como primeiro objetivo, comparar a mortalidade observada com os níveis de mortalidade preconizados pelos APR-DRG, APACHE II, SAPS II e com a escala preditiva de mortalidade do *Staging*,

Conforme abordado na discussão metodológica a forma mais correta de operacionalizar a comparação foi distribuir o índice global de gravidade fornecido pelo *Staging*, APACHE II SAPS II por uma escala ordinal em tudo igual à escala de mortalidade disponibilizada pelos APR-DRG.

A seguir optou-se numa primeira fase por separar os sistemas genéricos (APR-DRG e Disease Staging) dos sistemas clínicos (APACHE II e SAPS II) e fazer a comparação dois a dois e no fim comparar os sistemas genéricos com os sistemas clínicos.

No presente estudo é de relevante importância, porque confere significância clínica aos produtos definidos pelo *Staging*, APACHE II e SAPS II e a retira aos APR-



DRG (Costa, Lopes e Santana, 2008).

De resto, os resultados vêm corroborar as afirmações de Costa et al (2008), quando estes citando diversos autores afirmam que os DRGs não apresentam significância clínica e que não são facilmente interpretáveis e utilizáveis pelos médicos.

Mais, é possível observarem-se casos em que doenças principais com o mesmo estágio e subestadio de gravidade são codificadas em GDHs diferentes. Exemplo disso são os casos codificados pelo *Staging* como GIS8 – Outras doenças do esófago, estômago e duodeno. No seu estágio de gravidade 2,01, caracterizado por hemorragia, obstrução é possível encontrarem-se códigos de diagnóstico principais que vão desde o código ICD-9-CM 530 a 530 que são partilhados por dois GDHs diferentes, o GDH 220 - Procedimentos major no estomago, esófago e/ou duodeno e o GDH 253 Medico – Outras hemorragias gastrointestinais inespecíficas. Isto é, neste caso observaram-se episódios com os mesmos diagnósticos principais, aos quais o *Staging* atribuí a mesma doença com iguais estágios e subestadios de gravidade, mas os APR- DRG atribuem diferentes GDHs em função do tratamento eleito.

De resto, esta particularidade já havia sido identificada noutros estudos e leva a que os APR-DRG possuam neutralidade económica, isto é, considerem os tratamentos disponíveis para a definição de novos produtos.

Idealmente um sistema de classificação de doentes não deve possuir neutralidade económica, isto é não deve considerar os tratamentos disponíveis nem tão pouco as características de um determinado hospital. Ao fazê-lo está a considerar o risco institucional de um determinado doente ser admitido num determinado hospital ao invés de considerar somente o risco clínico intrínseco a cada doente. Costa, Lopes e Santana, 2008; Iezzoni et al., 1991, 1991; Schwartz et al., 1996).

## 6. Conclusões

Os APR-DRG e o Disease Staging (Coded) Disease Staging são retrospectivos, o que constitui desvantagem comparativamente aos modelos clínicos que disponibilizem informações sobre a situação do doente no momento da admissão, durante a estadia e na alta o que é mais adequada para a identificação e medição da produção hospitalar.

Os dois sistemas expõem vantagens e desvantagens para a administração hospitalar consoante a finalidade pretendida. Contudo, quando o pretendido é a realização de estudo da mortalidade é imperativo recorrer-se ao *Disease Staging*. (Costa 2008). Em concordância com o descrito anteriormente o presente estudo vem reiterar que o Disease Staging não só apresenta maior sensibilidade para identificar os doentes que efetivamente se espera que não venham a falecer, bem como aqueles que se espera que venham a falecer.

Para este facto contribuíram diversos fatores, mas o mais importante está na dependência da definição de produto hospitalar, própria de cada sistema. Portanto, para o estudo de mortalidade é fundamental um sistema como o Disease Staging, que agrupa os doentes em “*clusters*” que requerem tratamentos idênticos e têm resultados esperados idênticos e que intrinsecamente está capacitado para medir a gravidade da doença. As características descritas permitem o Disease Staging obter a significância clínica do sistema, imprescindível para facilitar a perceção não só aos gestores como aos clínicos, compreender e analisar quais as patologias com maior ou menor mortalidade, quais os fatores de risco que condicionam essa mortalidade bem como os tratamentos e complicações que influenciaram positivamente ou negativamente os resultados de saúde. Neste estudo ficou mais uma vez demonstrado que os APR-DRG além de não possuírem significância clínica possuem neutralidade económica.

Os APR-DRG não fazem distinção entre comorbilidade e complicação resultando daí, incorretamente a valorização do risco “oferecido” pelo hospital ao invés de considerar somente o risco intrínseco a cada doente (procura). Outro problema inerente aos APR-DRG é não atribuição de relevância necessária a idade como dimensão de risco, bem como de refletirem a demora média na atribuição dos níveis de mortalidade.

No que respeita às duas escalas disponibilizadas pelos APR-DRG há ainda a destacar um facto curioso, o superior desempenho da escala de severidade face à escala de mortalidade na previsão da mortalidade. Ainda nesta área há a salientar a utilidade de uma escala de razão como a do Disease Staging porque permite a comparação entre episódios e conhecer o risco de morte individual para cada um.

No que respeita aos sistemas de classificação de doentes clínicos, são prospetivos o que constitui uma considerável vantagem em relação aos sistemas de classificação de doentes genéricos. Quer o APACHE II quer o SAPS II são sistemas com operacionalização nas unidades de cuidados intensivos onde as decisões são a quente, ou seja, imediatamente após a admissão dos doentes.

As dimensões do risco utilizados nos sistemas clínicos permite-lhes adquirir a indispensável significância clínica, isto é, a característica que permite não só aos gestores como aos clínicos, compreender e analisar quais as patologias com maior ou menor mortalidade, quais os fatores de risco que condicionam essa mortalidade bem como os tratamentos e complicações que influenciaram positivamente ou negativamente os resultados de saúde

A dificuldade deparada foi os dois sistemas não disponibilizarem o risco de morte individual.

Os sistemas de classificação de doentes clínicos desenvolvidos mais recentemente como o APACHE III, APACHE IV e SAPS III já prevê resultado de doente individual, identifica os fatores que influenciam as variações de resultados dos cuidados e tempo de permanência nas UCIs (Knaus et al, 1993; Zimmerman et al 2006; Rothen et al, 2007; Moreno et al, 2005).

Em resumo, a mortalidade é utilizada como indicador de efetividade dos cuidados prestados e indicador de qualidade. Para desempenhar esta última função deve estar ajustada as características dos doentes como diferenças clínicas, demográficas e de *casemix* à admissão hospitalar.

A mortalidade hospitalar apresenta uma série de vantagens face a outros períodos de observação fora do hospital: fácil de obter; está dependente de informação após a alta; incluir todos os doentes independentemente da duração de internamento; capta o que efetivamente ocorreu dentro do episódio de internamento, sendo comparável com outros hospitais (Lopes 2010).

Segundo vários estudos realizados a mortalidade ocorrida até os 30 dias após a admissão tende a ser semelhante à mortalidade intra-hospitalar e que 30 dias após admissão tem pouco a ver com efeitos hospitalares específicos. (Häkkinen et al., 2015).

No presente estudo ficou patente, porque a finalidade para a qual foram desenvolvidos não são coincidentes. Os APR-DRG focam-se no consumo de recursos e como tal na dimensão ligadas à oferta de cuidados. Por sua vez o *Disease Staging* está focado nas características da doença e no risco de morte, portanto posiciona-se do lado da procura. Estas características como demonstrado, influenciam o desempenho dos sistemas no estudo da mortalidade.

Quanto aos sistemas clínicos de classificação de doentes seguindo os mesmos fundamentos do *Disease Staging* posicionam do lado da procura. Os doentes gravemente doentes são admitidos e tratados nas unidades de cuidados intensivos onde o índice de gravidade e o risco de morte dos doentes são determinados pelo APACHE II e o SAPS II.

O estudo apresenta limitações relacionadas com o (1) tamanho da amostra, (2) com os sistemas de classificação de doentes (3) e tipo e o modo como foram colhidos os dados.

A base de dados inicial incluía 2439 episódios de internamento e foram excluídos 852 episódios de internamento de acordo os critérios anteriormente referidos. Como resultado a base do desenvolvimento do estudo consistiu em 1587 episódios de internamento. O *Staging* e os APR-DRG foram desenvolvidos de grandes populações de doentes internados nos hospitais de agudos. Por sua vez o APACHE II e o SAPS II as populações de doentes não foram tão grandes como nos sistemas genéricos, mas tiveram a seguinte magnitude: APACHE II 5515 doentes e SAPS II 12997 doentes.

Ainda em relação aos sistemas de classificação de doentes clínicos são das primeiras gerações e não determinam risco de morte individual.

Em relação ao momento de colheita de dados os sistemas genéricos são retrospectivos e dependem da confiabilidade de codificação (Iezzoni, 2012). Os sistemas clínicos apresentam vantagens como prospetividade e confiabilidade. No entanto, a confiabilidade intra-avaliador e inter-avaliador pode não ser adequada porque estão restringidos às UCIs e não existe auditorias as colheitas e a avaliação da qualidade dos dados.

Em Portugal, para que os sistemas de classificação de doentes clínicos que avaliam a severidade do estado doente como, APACHE, SAPS, Disease Staging e APR-DRG possam ser utilizados de forma sistemática será necessário alterar a folha de admissão e alta e concomitantemente o resumo informatizado de alta, com inclusão de dados clínicos ou seja, elementos referentes a sintomas, sinais vitais e de disfunção fisiológica.

Se a evolução for no sentido da generalização destes sistemas de classificação de doentes poderá conduzir a sistemas mistos com ganho na previsão da morte, custos e duração de internamento.

Do estudo conclui-se que os sistemas com significância clínica, como Disease Staging, APACHE II e SAPS II têm maior sensibilidade na previsão da mortalidade e estão relacionados com a procura de cuidados de saúde. Neste contexto podem ser uma ferramenta importante para os gestores hospitalares ao permitir conhecer o tipo de doentes que tratam, os custos incorridos e as necessidades em recursos humanos e equipamento. Para a sociedade em geral pode ser um meio de conhecer os resultados para permitir uma escolha orientada. O financiamento adequado evita a seleção adversa e favorece a equidade

Os resultados demonstram que os doentes com nível de gravidade 1 e nível de gravidade 4 avaliado pelo APACHE II e O SAPS II tiveram internamento mais curtos. Num estudo anterior Naved e colaboradores (2011) também obtiveram resultados semelhantes. Os doentes com índice de gravidade mais elevada além de morrerem mais, têm internamentos mais curtos.

## 7. REFERÊNCIAS BIBLIOGRÁFICAS

3M™ APR DRG Classification System Version 31.0 (effective 10/01/2013). Methodology Overview

ACSS. Circular informativa Nº 24/2016/DPSACSS de 05/07/2016. Implementação do Sistema de codificação clínica ICD-10-CM/PCS Lisboa. Portugal.

ACSS. Circular normativa Nº 8/2013/DPS de 04/03/2013. Indicador PNA-Presente na Admissão.

Ahmed S, Berzon RA, Revicki DA, Lenderking WR, Moinpour CM, Basch E, Reeve BB, Wu AW; International Society for Quality of Life Research. The use of patient-reported outcomes (pro) within comparative effectiveness research: Implications for clinical practice and health care policy. *Med Care*. 2012; 50(12):1060-1070.

Appleby J, Harrison T, Hawkins L, Dixon A. Ideas that change health care. The King's Fund [Internet] 2012 [cited 2018 Oct]. Available from: [www.kingsfund.org.uk/publications](http://www.kingsfund.org.uk/publications).

Ash AA, Shwartz. M. Evaluating the Performance of Risk-Adjustment Methods: Dichotomous Variables. In: Iezzoni LI, editor. *Risk Adjustment for Measuring Healthcare Outcomes*. 2nd Ed. Chicago: Health Administration Press; 1997. p. 427-469.

Averill RF, Goldfield N, Hughes JS, Bonazelli J, McCullough EC, Mullin R et al. 3MTM APR DRG Classification System. Methodology Overview. Version 26.1 (effective 10/10/2008). 3M Health Information system.

Averill RF, Muldoon JH, Vertrees JC, Goldfield NI, Mullin RL, Fineran EC et al. The evaluation of case mix measurement using diagnosis related groups (DRG's). 3M Health Information Systems. Reports 5.1998. p. 1-40.

Averill RF, Muldoon JH, Vertrees JC, Goldfield NI, Mullin RL, Fineran EC, et al. The Evolution of Case mix Measurement Using Diagnosis Related Groups. 3M HIS Working Paper 10-97. Wallingford, CT, 1997.

Aylin P, Bottle A, Majeed A. Use of administrative data or clinical databases as predictors of risk of death in hospital: comparison of models *BMJ* [Internet].2007 [cited 2017 Jan]; 19;334(7602): 1044. Available from: Doi 10.1136/bmj.39168.496366.55

Baker DW, Gazmararian JA, Williams MV, Scott T, Parker RM, Green D, et al. Functional health literacy and the risk of hospital admission among Medicare managed care enrollees. *Am J Public Health*. 2002b;92(8):1278-83.

- Barer ML. Case mix adjustment in hospital cost analysis: information theory revisited. *Journal Health Economic*. 1982; 1(1):53-80.
- Barnum H, Kutzin J, Saxenian H. – Incentives and provider payment methods. *International Journal of Health Planning and Management*. 1995; 10: 23-45.
- Barros P, Gomes J. GANEC – Gabinete de Análise Económica Faculdade de Economia – Universidade Nova de Lisboa [Internet] 2002 [cited 2010 fev.] Available from: [http://www.tcontas.pt/pt/actos/rel\\_auditoria/2003/rel047-2003-estudo\\_externosumario.pdf](http://www.tcontas.pt/pt/actos/rel_auditoria/2003/rel047-2003-estudo_externosumario.pdf).
- Barros PP e Gomes JP. Os Sistemas Nacionais de Saúde da União Europeia e Principais Modelos de Gestão Hospitalar e Eficiência no Sistema Hospitalar Português. Lisboa: Faculdade de Economia, Universidade Nova de Lisboa; 2002.
- Barton ML Heckscher RA. Racial and ethnic identification practices in public health data systems in New England. *Public Health Rep*. 2002; 117(1): 50–61
- Baker DW, Gazmararian JA, Williams MV, Scott T, Parker RM, Green D et al. Functional health literacy and the risk of hospital admission among Medicare managed care enrollees. *Am J Public Health*. 2002 Aug;92(8):1278-83.
- Beck DH, Smith GB, Pappachan JV, Millar B. External validation of the SAPS II, APACHE II and APACHE III prognostic models in South England: a multicenter study. *Intensive Care Medicine*. 2003; 29: 249-256.
- Becker RB, Zimmerman JE, Knaus WA, Wagner DP, Seneff MG, et al. The use of APACHE III to evaluate ICU length of stay, resource use, and mortality after coronary artery by-pass surgery. *J Cardiovasc Surg*. 1995; 36 (1): 1–11.
- Bensley RP, Yoshida S, Lo RC, Fokkema M, Hamdan AD, Wyers MC, et al. Accuracy of administrative data versus clinical data to evaluate carotid endarterectomy and carotid stenting. *J Vasc Surg*. 2013;58(2):412-9.
- Bentes M, Berardo A: 8º Encontro Nacional de Economia da Saúde; 23 e 24 de outubro de 2003.
- Bentes M, Gonçalves M, Tranqueada S, Urbano J. A Utilização de GDH's como Instrumento de Financiamento Hospitalar. *Gestão Hospitalar*. 1996; 33 (9), 33-40.
- Bentes, M. – O financiamento dos hospitais. Lisboa: IGIF, 1998.
- Bentes M, Urbano J, Carvalho M, Tranquada S. — Using DRGs to fund hospitals in Portugal: an evaluation of the experience. Lisboa: Ministério da Saúde, 1991 (SIGSS Program, April 1991)

Ben-Tovim D, Woodman R, Harrison JE, Pointer S, Hakendorf P, Henley G. Measuring and reporting mortality in hospital patients. Australian Institute of Health and Welfare Canberra., March 2009.

Best WR, Cowper DC. The ratio of observed-to-expected mortality as a quality of care indicator in non-surgical VA patients. *Med Care* 1994; 32(4): 390-400.

Blumberg MS. Biased Estimates of Expected Acute Myocardial Infarction Mortality Using Medis-Groups Admission Severity Groups. *Journal of the American Medical Association*. 1991;265(22):2965-2970.

Blumberg MS. Comments on HCFA Hospital Death Rate Statistical Outliers. *Health Services Research*, 1987; 21 (6), 715-739.

Booth FV, Short M, Shorr, Arkins, Bates Qualy RL, Levy H. Application of a population-based severity scoring system to individual patients results in frequent misclassification. *Critical Care*20059:R522. Doi: 10.1186/cc3790. [Consultado a 01-04-2017]. Disponível em: <http://ccforum.com/content/9/5/R522>.

Bouch DC, Thompson JP. Severity scoring systems in the critically ill. *Critical Care & Pain*. 2008;8 181-185.

Brailer DJ, Kroch E, Pauly MV, Huang J. Comorbidity-adjusted complication risk: a new outcome quality measure. *Med Care*. 1996;34(5):490-505.

Breslow MJ, Badawi O. Severity scoring in the critically ill: part 1-interpretation and accuracy of outcome prediction scoring systems. *Chest*. 2012a;141(1):245-52.

Breslow MJ, Badawi O. Severity scoring in the critically ill: part 2: maximizing value from outcome prediction scoring systems. *Chest* 2012b; 141(2):518-527.

Brewster AC, Jacobs CM, Bradbury RC. Classifying severity of illness by using clinical findings. *Health Care Financ Rev*. 1984; (Suppl): 107–108.

Brook RH, Iezzoni LI, Jencks SF, Knaus WA, Krakauer H, Lohr KN et al. Symposium: Case-Mix Measurement and Assessing Quality of Hospital Care. *Health Care Financing Review*, December (Special Number). 1987:39-48.

Brook RH, McGlynn EA, Cleary PD. Quality of Health Care. Part 2: Measuring Quality of Health Care. *New England Journal of Medicine*.1996; 335: 966-970.

Bueno H, Ross JS, Wang Y, Chen J, Vidan MT, Normand SL, et al. Trends in length of stay and short-term outcomes among Medicare patients hospitalized for heart failure,



1993–2006. JAMA: the journal of the American Medical Association. 2010; 303(21):2141–7.

Cabral F. Comportamentos de Admissão de Doentes nos hospitais em Portugal: seleção adversa? Lisboa: Escola Nacional de Saúde Pública. Universidade Nova de Lisboa, 2004. Dissertação elaborada no âmbito do XXXII Curso de Especialização em Administração Hospitalar 2002/ 2004 ministrado pela ENSP/UNL.

Cassel CK, Vladeck BC. ICD-9 code for palliative or terminal care. New England Journal of Medicine. 1996;335(16):1232-4.

Castella X, Artigas A, Bion J, Kari A. A comparison of severity of illness scoring systems for intensive care unit patients: results of a multicenter, multinational study. The European/North American Severity Study Group. Critical Care Medicine. 1995; (23): 1327-1335.

Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: Development and validation. Journal of Chronic Disease 1987; 40:373.

Chassin MR, Loeb JM, Schmaltz SP, Wachter RM. Accountability measures—using measurement to promote quality improvement. N Engl J Med. 2010;363(7):683–688

Chen LM, Staiger DO, Birkmeyer, Ryan AM, Zhang W, Dimick JB. Composite Quality Measures for Common Inpatient Medical Conditions. Med Care. 2013; 51(9): 832–837.

Chong WF, Ding YY, Heng BH. A comparison of comorbidities obtained from hospital administrative data and medical charts in older patients with pneumonia. BMC Health Services Research 2011, 11:105. Doi: 10.1186/1472-6963-11-105. [Consultado a 17/08/2016]. Disponível em: <http://www.biomedcentral.com/1472-6963/11/105>.

Christensen S, Johansen MB, Christiansen CF, Jensen R, Lemeshow S. Comparison of Charlson comorbidity index with SAPS and APACHE scores for prediction of mortality following intensive care. Clinical Epidemiology. 2011;3:203-11.

Cody R. Cody's Data Cleaning Techniques Using SAS, Second Edition. Cary, NC: SAS Press. 2008

Coffey RM, Goldfarb MG. DRGs and disease staging for reimbursing Medicare patients. Medical Care. 1986;24(9):814-29.

Coffey, R., M. Milenkovic, and R. M. Andrews. The Case for the Present on Admission (POA) Indicator. Rockville, MD: U.S. Agency for Healthcare Research and Quality, 2006

Cohen ME, Bilimoria KY, Ko CY, Richards K, Hall BL. Effect of subjective preoperative variables on risk-adjusted assessment of hospital morbidity and mortality. *Annals Surgery*. 2009 Apr;249(4):682-9.

Cohen, Jacob. "A coefficient of agreement for nominal scales". *Educational and Psychological Measurement*. 1960;20 (1): 37–46

Coleman EA, Background Paper on Transitional Care Performance Measurement. Appendix I. Institute of Medicine, Performance Measurement: Accelerating Improvement. Washington, DC: National Academy Press; 2005.

Coleman, EA, Parry C, Chalmers S, Min SJ. The care transitions intervention: results of a randomized controlled trial. *Arch Intern Med*. 2006; 166:1822–1828..

Conklin JE, Lieberman JV, Barnes CA, Louis DZ. Disease staging: Implications for hospital reimbursement and management. *Health Care Financing Review Supplement* 1984; 13-22.

Cook R, Cook D, Tilley J, Lee K, Marshall J: Multiple organ dysfunction: baseline and serial component scores. *Critical Care Medicine* 2001; 29:2046-205

Costa C, Reis V. O sucesso nas organizações de saúde: *Revista Portuguesa de Saúde Pública*. 11:3 (1993), 59-68.

Costa C, Delgado M, Carvalho R. Relação entre as características da procura e a utilização de recursos num hospital. *Revista Portuguesa de Saúde Pública* 7(4) (1989) 27-40.

Costa C, Lopes S, Santana R. Diagnosis Related Groups e Disease Staging: importância para administração hospitalar. *Revista Portuguesa de Saúde Pública*. 7 (2008) 7-28

Costa C, Lopes S. Avaliação do desempenho dos hospitais SA. Lisboa: Grupo de Disciplinas de Gestão em Organização de Saúde. Escola Nacional de Saúde Pública. Universidade Nova de Lisboa, 2005.

Costa C, Lopes S. – Produção hospitalar: a importância da complexidade e da gravidade. *Revista Portuguesa de Saúde Pública*. 4 (2004) 35-50.

Costa C, Nogueira P. A medição da produção do hospital: a importância da fiabilidade. *Revista Portuguesa de Saúde Pública*. 12(2) (1994), 31-40.

Costa C, Santana R, Boto P. – Financiamento por capitação ajustada pelo risco: conceptualização e aplicação. *Revista Portuguesa de Saúde Pública*. 7 (2008) 67- 102.

Costa C, Santana R, Lopes S, Barriga N. A importância do apuramento de custos por doente: metodologias de estimação aplicadas ao internamento hospitalar português. *Revista Portuguesa de Saúde Pública* 7 (2008) 131-146.

Costa C, Silva S. Avaliação do desempenho dos hospitais públicos (Internamento) em Portugal Continental - versão provisória - 2013. Grupo de Disciplinas de Gestão em Organizações de Saúde. Escola Nacional de Saúde Pública Universidade Nova de Lisboa, Lisboa 2014.

Costa C. A severidade da doença – Identificação e caracterização de alguns sistemas de classificação. *Revista Portuguesa de Saúde Pública*, 9:1 (1991) 37-43.

Costa C. Ajustamento pelo risco: da conceptualização à operacionalização. *Revista da Escola Nacional de Saúde Pública*. 5, (2005) 7-38.

Costa C. Financiamento de Serviços de Saúde – A Definição de Preços. *Revista Portuguesa de Saúde Pública*. 8 (2), (1990) 65-72.

Costa C. Os DRGs (Diagnosis Related Groups) e a gestão do hospital. *Revista Portuguesa de gestão*. III/IV 8: 1 (1994) 47-65.

Costa C. Produção e Desempenho Hospitalar. Aplicação ao Internamento. Tese de candidatura ao grau de Doutor em Saúde Pública na especialidade de Administração de Saúde pela Universidade Nova de Lisboa. Escola Nacional de Saúde Pública Universidade Nova de Lisboa, 2005.

Daley J, Iezzoni LI, Schwartz M. Conceptual and practical issues in developing risk adjustment methods. In: Iezzoni LI. *Risk adjustment for measuring health care outcomes*, 3rd edn. Chicago: Health Administration Press, 2003, 179–206.

Daley J, Schwartz M. Developing Risk-Adjustment Methods. En: Iezzoni LI, editor. *Risk adjustment for measuring healthcare outcomes*. 2nd ed. Chicago, Illinois: Health Administration Press;1997:279-329

Daley J, Jencks S, Draper D, Lenhart G, Thomas N, Walker J. Predicting hospital-associated mortality for Medicare patients. A method for patients with stroke, pneumonia, acute myocardial infarction, and congestive heart failure. *JAMA*. 1988 Dec 23-30;260(24):3617-24.

Danis M, Garrett J, Harris R, Patrick DL. Stability of Choices about Life-Sustaining Treatments. *Annals of Internal Medicine*. 1994;120(7):567-573

Davis RB, Iezzoni LI, Phillips RS, Reiley P, Coffman GA, Safran C. Predicting in-hospital mortality. The importance of functional status information. *Medical Care*. 1995;33(9):906-21.

De Coster C, Quan H, Finlayson A, Gao M, Halfon P, Humphries KH, et al. Identifying priorities in methodological research using ICD-9-CM and ICD-10 administrative data: report from an international consortium. *BMC Health Services Research* 2006, 6:77. doi:10.1186/1472-6963-6-77. [Consultado a 01-12-2016].

Demir E, Chausalet TJ, Xie H, Millard PH. Emergency readmission criterion: a technique for determining the emergency readmission time window. *IEEE Trans Inf Technol Biomed*. 2008; 12(5):644–9.

Des Harnais S, McMahon LF Jr, Wroblewski R. Measuring outcomes of hospital care using multiple risk-adjusted indexes. *Health Serv Res*. 1991;26(4):425-45.

Des Harnais SI, Forthman MT, Lowry JM, Wooster LD. Risk-Adjusted Clinical Quality Indicators: Indexes for Measuring and Monitoring Rates of Mortality, Complications, and Readmissions. *Quality Management in Health Care*. 2000; 9 (1):14-22.

Deyo RA, Cherkin DC, Ciol MA. Adapting a clinical comorbidity index for use with ICD-9-CM administrative databases. *Journal of Clinical Epidemiology*. 1992;45(6):613-9.

Diário da República n.º 153/2015. Portugal. Ministério da Saúde. Administração Central do Serviço de Saúde – Hospitais SNS: Termos de Referência para contratualização hospitalar no SNS – Contrato-Programa 2016.

Dimick JB, Welch HG, Birkmeyer JD. Surgical mortality as an indicator of hospital quality: the problem with small sample size. *JAMA*. 2004;292:847-851.

Direcção Geral de Saúde. Desempenho Comparado das Unidades de Saúde do SNS – Indicadores Mensais, 2002. DGS, Lisboa, 2002.

Donabedian A. Evaluating the quality of medical care. *Milbank Q*. 2005; 83(4): 691–729.

Donabedian A. Methods for deriving criteria for assessing the quality of medical care. *Medical Care Review*. 1980; 37(7):653-698.

Donabedian A. The epidemiology of quality. *Inquiry*. 1985; 22(3):282-292.

Donabedian A. The quality of care. How can it be assessed? *JAMA* 1988; 260(12): 1743–1748.

Donabedian A. Explorations in Quality Assessment and Monitoring: The definition of quality and approaches to its assessment. Health Administration Press, 1980

Efron B. Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics* 1979; 7 (1): 1-26

Elixhauser A, Steiner C, Harris DR, Coffey RM. Comorbidity measures for use with administrative data. *Medical Care*. 1998;36(1):8-27.

Elixhauser Comorbidity Software (última versão 3.7). Healthcare Cost and Utilization Project (UCUP). Agency for Healthcare Research and Quality, USA. [Consultado a 05/11/2016]. Disponível em: [www.hcup-us.ahrq.gov/sidoverview.jsp](http://www.hcup-us.ahrq.gov/sidoverview.jsp) .

Escobar GJ, Greene JD, Scheirer P, Gardner MN, Draper D, Kipnis P. Risk-adjusting hospital inpatient mortality using automated inpatient, outpatient, and laboratory databases. *Medical Care*. 2008;46(3):232-9.

Escoval A. – Sistemas de financiamento da saúde: análise e tendências. Dissertação para obtenção do Grau de Mestre em Ciências Empresariais. Instituto Superior de Ciências do Trabalho e da Empresa, Lisboa 1999

Evers SM, Struijs JN, Ament AJ, van Genugten ML, Jager JH, van den Bos GA. International comparison of stroke cost studies. *Stroke* 2004;35(5): 1209- 1215.

Fava GA, Tomba E, Sonino N. Clinimetrics: the science of clinical measurements. *International Journal of Clinical Practice*. 2012;66(1):11-5.

Fairclough DL. Patient reported outcomes as endpoints in medical research. *Stat Methods. Med Res*. 2004; 13(2):115–138.

Ferreira FL, Bota DP, Bross A, Mélot C, Vincent JL Serial evaluation of the SOFA score to predict outcome in critically ill patients. *JAMA* 2001; 286: 1754- 1758.

Fetter RB, Shin Y, Freeman JL, Averill RF, Thompson JD. Case mix definition by diagnosis-related groups. *Med Care* 1980;(suppl) 18:1–53

Fischer C, Anema HA, Klazinga NS. The validity of indicators for assessing quality of care: a review of the European literature on hospital readmission rate. *Eur J Public Health*. 2012; 22(4):484–91.

Flora JD. A method for comparing survival of burn patients to a standard survival curve. *J Trauma*. 1978;18(10):701-5.

Fokkema M, Hurks R, Curran T, Bensley RP, Hamdan AD, Wyers MC, et al. The impact of the present on admission indicator on the accuracy of administrative data for carotid endarterectomy and stenting. *J Vasc Surg*. 2014 Jan;59(1):32-8.e1. Doi: 10.1016/j.jvs.2013.07.006. Epub 2013 Aug 28. [Consultado a 10/11/2015].

Forthman MT, Gold RS, Dove HG, Henderson RD. Risk-adjusted indices for measuring the quality of inpatient care. *Qual Manag Health Care*. 2010;19(3):265-77.

Freitas A, Silva-Costa T, Lopes F, Garcia-Lema I, Teixeira-Pinto A, Brazdil P et al. Factors influencing hospital high length of stay outliers. *BMC Health Serv Res*. 2012;12:265. <http://www.biomedcentral.com/1472-6963/12/265> [Consultado a 12/07/2016].

Gagne JJ, Glynn RJ, Avorn J, Levin R, Schneeweiss S. A combined comorbidity score predicted mortality in elderly patients better than existing scores. *Journal of Clinical Epidemiology* 2011; 64(7): 749–75

Gago ACB. O Financiamento Hospitalar e os Grupos de Diagnóstico Homogêneos. Universidade de Aveiro. 2008

Gardner MJ, Altman DG. Statistics with confidence. *British Medical Journal*, London, 1989.

Geraci JM, Ashton CM, Kuykendall DH, Johnson ML, Wu L. International Classification of Diseases, 9th Revision, Clinical Modification: codes in discharge abstracts are poor measures of complication occurrence in medical inpatients. *Med Care*. 1997;35(6):589-602.

Geraci JM. In-hospital complication occurrence as a screen for quality-of-care problems: what's next? *Med Care*. 2000;38(8):777-80.

Geraci JM. The Demise of Comparative Provider Complication Rates Derived from ICD-9-CM Code Diagnoses. *Medical Care*, 2002; 40 (10), 847-850.

Ghanem-Zoubi NO, Vardi M, Laor A, Weber G, Bitterman H. Assessment of disease-severity scoring systems for patients with sepsis in general internal medicine departments. *Critical Care* 2001;15:R95. Doi: 10.1186/cc10102. [Consultado a 11/11/2017]. Disponível em: <http://ccforum.com/content/15/2/R95>

Glance LG, Dick AW, Osler TM, Mukamel DB. Does date stamping ICD-9-CM codes increase the value of clinical information in administrative data? *Health Serv Res*. 2006;41:231–251.

Glance LG, Li Y, Osler TM, Dick A, Mukamel DB. Impact of patient volume on the mortality rate of adult intensive care unit patients. *Crit Care Med* 2006; 34:1925–1934

Glance LG, Li Y, Osler TM, Mukamel DB, Dick AW. “Impact of the Present-on-Admission Code in Administrative Data on Patient Safety Measurement in Patients Undergoing CABG: Experience with the AHRQ Patient Safety Indicators.” *BMC Health Services*

Research. 2008; 8 (1): 176. Doi: 10.1186/1472-6963-8-176. [Consultado a 13/09/2018]. Disponível em: <http://www.biomedcentral.com/1472-6963/8/176>

Goldman LE, Chu PW, Bacchetti P, Kruger J, Bindman A. Effect of Present-on-Admission (POA). Reporting Accuracy on Hospital Performance Assessments Using Risk-Adjusted Mortality. *Health Serv Res.* 2015;50(3):922-38. Doi: 10.1111/1475-6773.12239. [Consultado a 13/09/2018].

Gonnella JS. Patient Case Mix: Implications for Medical Education and Hospital Costs. *J of Medical Education.* 1981; 56: 610 – 11.

Gonnella JS, Hornbrook MC, Louis DZ. Staging of disease: a case-mix measurement. *JAMA.* 1984; 251 (5): 637-644.

Gonnella JS, Louis DZ. Evaluation of ambulatory care. *Journal of Ambulatory Care Management.* 1988; 11: 68-83.

Gonnella JS, Louis DZ. Physicians responsibilities and the evaluation of outcomes of medical care. In Markson, L. M., Nash, D. B., eds. — *Accountability and Quality in Health Care: The New Responsibility.* Oakbrook, IL: Joint Commission on Accreditation of Healthcare Organizations, 1995.

Gonnella JS, Louis DZ. Severity of illness in the assessment of quality: disease staging. In Hughes Edward, F. X., ed. — *Perspectives on Quality in American Health Care.* Washington, DC: McGraw-Hill, 1988.

Gonnella JS, Louis DZ, Gozum ME, Callahan CA, Barnes CA, eds — *Clinical criteria for disease staging.* 5th edition. Ann Arbor, Michigan: Thomson-MEDSTAT, 2003.

Gonnella JS, Louis DZ, Gozum MVE, Callahan CA, Barnes CA, eds. *Disease Staging: Clinical and Coded Criteria.* Version 5.27. Thomson Reuters 2010.

Gonnella JS, Louis DZ, Gozum MVE, Callahan CA, Barnes CA, eds. *Disease Staging: Clinical Criteria,* 6th ed. Ann Arbor, MI: Truven Health Analytics, 2012.

Gonnella JS, Louis DZ, Gozum MVE, Callahan CA, Cathleen A. Barnes CA. *Disease Staging Clinical and Coded Criteria,* Version 5.26. 26 Janeiro 2009.

Gonnella JS, Louis DZ, Zeleznik C, Turner BJ. The problem of late hospitalization: a quality and cost issue. *Academic Medicine.* 1990; 65 (5): 314-319.

Gonnella JS, Louis DZ. Severity of illness and evaluation of hospital performance. *Revista Portuguesa de Saúde Pública.* 2005;23: 39-46.

Gonnella JS, Louis DZ. Severity of illness in the assessment of quality: Disease Staging. In Hughes Edward, F. X., ed. — Perspectives on Quality in American Health Care. Washington, DC: McGraw-Hill, 1988.

Goodacre S, Campbell M, Carter A. What do hospital mortality rates tell us about quality of care? *Emergency Medicine Journal*. 2015;32(3):244-7.

Green J, Passman LJ, Wintfeld N. — Analyzing Hospital Mortality: The Consequences of Diversity in Patient Mix. *JAMA*. 1991; 265 (14), 1849-1853.

Griffith J R, Alexander JA, Jelinek RC. Measuring Comparative Hospital Performance. *Journal of Healthcare Management*. 2002; 47 (1), 41-56.

Gruper P, Jaques A. Organization and management Professionalism. Update 2012. *Critical Care Medicine*. London

Guru V, Tu JV, Etchells E, Anderson GM, Naylor CD, Novick RJ, et al. Relationship between preventability of death after coronary artery bypass graft surgery and all-cause risk-adjusted mortality rates. *Circulation* 2008; 117(23): 2969-76.

Guru V, Fremes SE, Naylor CD, Austin PC, Shrive FM, Ghali WA, et al. Public versus private institutional performance reporting: What is mandatory for quality improvement? *Am Heart J*. 2006;152:573-8.

Guttmacher AE, Collins FE. "Genomic Medicine—A Primer." *New England Journal of Medicine*. 2002; 347 (19): 1512-20.

Häkkinen U, Rosenqvist G, Iversen T, Rehnberg C, Timo T. Seppälä, on behalf of the Euro HOPE study group. Outcome, Use of Resources and Their Relationship in the Treatment of AML, Stroke and Hip Fracture at European Hospitals. *Health Economic*. Volume 24, Issue Supplement S2 December 2015: 116–139

Halfon P, Egli Y, Pretre-Rohrbach I, Meylan D, Marzzi A, Burnand B. Validation of the Potentially Avoidable Hospital Readmission Rate as a Routine Indicator of quality of Care. *Medical Care* 2006; 44 (11), 972-981.

Hanchate A, Kronman AC, Young-Xu Y, Ash AS, Emanuel E. Racial and ethnic differences in end-of-life costs: why do minorities cost more than whites? *Archives of Internal Medicine*. 2009;169(5):493-501.

Hanley J, McNeil B. Meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982; 143: 29-36.

Hannan EL, Kilburn H Jr, Lindsey ML, Lewis R. Clinical versus administrative data bases for CABG surgery: does it matter? *Med Care*. 1992; 30:892–907.



Hannan EL, Racz MJ, Jollis JG, Peterson ED. Using Medicare claims data to assess provider quality for CABG surgery: does it work well enough? *Health Serv Res.* 1997;31:659-678.

Harrell FE Jr, Lee KL, Pollock BG. Regression models in clinical studies: determining relationships between predictors and response. *Institute* 1988;80(15):1198-202.

Harrel FA Jr, Lee KL, Califf RM, Pryor DB, Rosati RA. Regression Modelling Strategies for Improved Prognostic Prediction. *Statistics Medicine.* 1984;3(2):143-52.

Harrell FE Jr, Lee KL, Mark DB. Tutorial in biostatistics. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med.* 1996;15:361-87.

Healthcare Cost and Utilization Project (HCUP) Methods Series Report on The Case for the POA Indicator. Report 2015-06

Heggestal T, Lillengs. *International Journal for Quality in Health Care.* 2003;15(2): 147-154

Heijink R, Koolman X, Pieter D, van der Veen A, Jarman B, Gert Westert G. Measuring and explaining mortality in Dutch hospitals; The Hospital Standardized Mortality Rate between 2003 and 2005. *BMC Health Services Research* 2008; 8:73. Doi:10.1186/1472-6963-8-73. [Consultado a 07/01/2017]. Disponível em: <http://www.biomedcentral.com/1472-6963/8/73>.

Hightower CE, Riedel BJ, Feig BW, Morris GS, Ensor JE Jr, Woodruff VD et al. A pilot study evaluating predictors of postoperative outcomes after major abdominal surgery: Physiological capacity compared with the ASA physical status classification system. *British Journal of Anaesthesia.* 2010;104(4):465-71.

Hill CA, Winfrey KL, Rudolph BA. "Best hospitals": a description of the methodology for the Index of Hospital Quality. *Inquiry.* 1997 Spring;34(1):80-90

Hofer TP, Hayward RA. Identifying poor-quality hospitals. Can hospital mortality rates detect quality problems for medical diagnoses? *Med Care* 1996;34:737-53

Holloway RG, Quill TE. Mortality as a measure of quality: implications for palliative and end-of-life care. *Journal of the American Medical Association.* 2007;298(7):802-4.

Holmes J, Powell-Griner E, Lethbridge-Cejku M, Heyman K. Aging Differently: Physical Limitations Among Adults Aged 50 years and Over: United States, 2001–2007. *NCHS Data Brief.* 2009;(20):1-8.

Horn SD, Horn RA, Sharkey PD. The Severity of Illness Index as a severity adjustment to diagnosis-related groups. *Health Care Financing Review*. 1984; Suppl:33-45.

Horn SD, Sharkey PD, Chambers AF, Horn RA. Severity of illness within DRGs: impact on prospective payment. *Am J Public Health*. 1985;75(10): 1195–1199

Horn DS, Horn RA. The Computerized Severity Index. *Journal of Medical Systems*. 1986; 10 (1):73–78

Hornbrook M, Goodman M. Chronic disease, functional health status, and demographics: a multi-dimensional approach to risk adjustment. *Health Services Research*. 1996;31(3): 283-307.

Hornbrook MC. Hospital case mix: its definition, measurement and use: Part I. The conceptual framework. *Med Care Rev* 1982; 39:1-72.

Hornbrook MC. Hospital case mix: its definition, measurement and use: Part II. Review of alternative measure. *Med Care Rev* 1982; 39 (2):72-123

Hornbrook MC. Techniques for Assessing Hospital Case Mix, *Annual Review of Public Health*. 1985;6: 295-322.

Hornbrook MC, Hurtado AV, Johnson RE. Health care episodes: definition, measurement and use. *Medical Care Review*. 1985 Fall;42(2):163-218.

Hosmer DW, Lemeshow S. *Applied logistic regression*. 2nd edn. 2000. New York: John Wiley & Sons Inc.

Hughes JS, Averill RF, Goldfield NI, Gay JC, Muldoon J, McCullough E et al. Identifying potentially preventable complications using a present on admission indicator. *Health Care Financ Rev*. 2006 Spring;27(3):63-82. Disponible en: <http://www.semicyuc.org/temas/calidad/indicadores-de-calidad>.

Hughes JS, Iezzoni LI, Daley J, Greenberg L. How severity measures rate hospitalized patients. *J Gen Intern Med*. 1996;11(5):303-11

De Coster C, Quan H, Finlayson A, Gao M, Halfon P, Humphries KH et al. Identifying priorities in methodological research using ICD-9-CM and ICD-10 administrative data: report from an international consortium. *BMC Health Services Research* 2006, 6:77. Doi:10.1186/1472-6963-6-77. [Consultado a 01-12-2016].

Iezzoni L. *Risk Adjustment for Measuring Healthcare Outcomes*. 4 th Edition. 2012.

Iezzoni L. *Risk Adjustment for Measuring Healthcare Outcomes*. Health Administration Press, Ann Arbor, MI. 1994.

Iezzoni L, Ash AS, Shwartz M, Dale J, Hughes JS, Mackiernan YD. Judging Hospitals by Severity-Adjusted Mortality Rates: The Influence of the Severity-Adjustment Method. *American Journal of Public Health*.1996b; 86 (10): 1379- 1387

Iezzoni L. Risk Adjustment for Medical Outcome Studies. In Grady ML, Schwartz HA (edt.) *Medical Effectiveness Research Data Methods*. Agency for Health Policy and Research. US Department of Health and Human Services. Rockville 1992a.

Iezzoni L. The Risks of Risk Adjustment. *Journal of the American Medical Association*, 1997a; 278 (19), 1600-7.

Iezzoni LI Case classification and quality of care, Issues to consider before making the investment. *Qual Rev Bull*, 1987; 13:135-9.

Iezzoni LI, Daley J, Heeren T, Foley SM, Fisher ES, Duncan C, Hughes JS, Coffman GA. Identifying complications of care using administrative data. *Medical Care*. 1994;32(7):700-715.

Iezzoni LI, Ash AS, Coffman GA, Moskowitz MA. Predicting In-Hospital Mortality – A Comparison of Severity Measurement Approaches. *Medical Care*, 1992b; 30 (4), 347-359.

Iezzoni LI, Ash AS, Shwartz M, Mackiernan YD. Differences in procedure use, in-hospital mortality, and illness severity by gender for acute myocardial infarction patients: are answers affected by data source and severity measure? *Med Care*. 1997e;35:158-171

Iezzoni LI, Ash AS, Shwartz M, Landon BE, Mackiernan YD. Predicting in-hospital deaths from coronary artery bypass graft surgery: do different severity measures give different predictions. *Med Care*. 1998;36: 28-39.

Iezzoni LI, Restuccia JD, Shwartz M, Schaumburg D, Coffman GA, Kreger BE et al. The Utility of Severity of Illness Information in Assessing the Quality of Hospital Care – The Role of Clinical Trajectory. *Medical Care*, 1992c; 30 (5), 428-444.

Iezzoni LI, Shwartz M, Ash AS, Mackiernan YD. Predicting in-hospital mortality for stroke patients: results differ across severity-measurement methods. *Med Decis Making*.1996a;16(4):348-356.

Iezzoni LI, Shwartz M, Ash AS, Hughes JS, Daley J, Mackiernan YD. Severity measurement methods and judging hospital death rates for pneumonia. *Medical Care*. 1996c Jan;34(1):11-28

Iezzoni LI, Shwartz M, Ash AS, Mackiernan YD. Using severity measures to predict the likelihood of death for pneumonia inpatients. *Journal of General Internal Medicine*. 1996d ;11(1):23-31

Iezzoni LI. Finally present on admission but needs attention. *Medical Care*. 2007; 45(4): 280- 282.

Iezzoni LI. Risk Adjustment for Measuring Healthcare Outcomes. Ann Arbor, Mich: Health Administration Press. Chicago; 1994.

Iezzoni LI. Statistically Derived Predictive Models – Caveat Emptor. *Journal of General Internal Medicine*, 1999;14 (6), 388-389.

Iezzoni LI. Using administrative diagnostic data to assess the quality of hospital care: pitfalls and potential of ICD-9-CM, *International Journal of Technology Assessment in Health Care*. 1990; 6(2), 272 – 281.

Iezzoni, LI. Dimensions of Risk. In Iezzoni LI (editor) — «Risk Adjustment for Measuring Healthcare Outcomes». 2nd Edition. Chicago: Health Administration Press. 1997b; 43-167.

Iezzoni LI, Ash AA, Shwartz M, Daley J, Hughes J, MacKiernan YD. Predicting Who Dies Depends on How Severity is Measured: Implications for Evaluating Patient Outcomes. *Annals of Internal Medicine*. 1995a; 123 (10), 763-770.

Iezzoni LI, Shwartz M, Ash AA, Hughes JS, Daley J, Mackiernan YD. Using Severity-Adjusted Stroke Mortality Rates to Judge Hospitals. *International Journal for Quality in Health Care*. 1995b; 7 (2), 81-94.

Iezzoni LI. Assessing quality using administrative data. *Annals of Internal Medicine*. 1997c;127: 8 Part 2:666-674.

Iezzoni LI. Risk adjustment for measuring health care outcomes. 3rd ed. Chicago, IL: Health Administration Press, 2003.

Iezzoni LI, Greenberg, MS. Capturing and Classifying Functional Status Information in Administrative Databases. *Health Care Financ Rev*. 2003 Spring; 24(3): 61–76.

Iezzoni LI. Risk and outcomes. In L.I. Iezzoni (Ed.), *Risk adjustment for measuring healthcare outcomes* (pp. 1–41). Chicago: Health Administration Press, 1997d.

Iezzoni LI. Coded data from administrative sources. In: Iezzoni, LI., editor. *Risk adjustment for measuring health care outcomes*. 4. Chicago, Illinois: Health Administration Press; 2013a. p. 95-146.

Iezzoni LI. Reasons for risk adjustment. In: Iezzoni, LI., editor. Risk adjustment for measuring health care outcomes. 4. Chicago, Illinois: Health Administration Press; 2013b. p. 1-14.

Iezzoni LI, Davis RB, Palmer RH, Cahalane, M, Hamel MB, Mukamal K et al. Does the Complications Screening Program Flag Cases with Process of Care Problems? Using Explicit Criteria to Judge Processes. *Int J Qual Health Care*.1999; 11 (2), 107-118

Iezzoni LI, Daley J. A description and clinical assessment of the Computerized Severity Index. *QRB Quality Review Bulletin*. 1992 Feb;18(2):44-52.

Jencks SF, Daley J, Draper D, Thomas N, Lenhart G, Walker J. Interpreting Hospital Mortality Data – The Role of Clinical Risk Adjustment. *JAMA*, 1998; 260 (24), 3611-3616.

Jencks SF, Williams MV, Coleman EA. Rehospitalizations among patients in the Medicare fee-for-service program. *New England Journal Medicine*. 2009;360(14):1418-28.

Johnson R, Hornbrook M, Nichols G. Replicating the chronic disease score from automated pharmacy data. *Journal of Clinical Epidemiology*.1994; 47,10: 1191-1199.

Joynt KE, Jha AK. Thirty-day readmissions — truth and consequences. *New England Journal Medicine*. 2012; 366(15):1366–1369.

Junger A, Engel J, Benson M, Bottger S, Grabow C, Hartmann B et al. Discriminative power on mortality of a modified Sequential Organ Failure Assessment score for complete automatic computation in an operative intensive care unit. *Critical Care Medicine*. 2002;30(2):338-42.

Jutice AA, covinsky KE, Berlin JA. Assessing the generalizability of prognostic information. *Ann. Intern. Med.*, 1999; 130: 515-524.

Kaga K, Gonnella JS. Disease Staging. *Japanese Journal of Nursing Education*. 1990; 31: 595-598.

Keegan MT, Bekele Afessa B. Severity of illness scoring systems in the intensive care unit, *Critical Care Medicine* 2010; 38(12).

Khuri SF, Daley J, Henderson W, Hur K, Demakis J, Aust JB et al. The Department of Veterans Affairs' NSQIP: the first national, validated, outcome-based, risk-adjusted, and peer-controlled program for the measurement and enhancement of the quality of surgical care. National VA Surgical Quality Improvement Program. *Annals of Surgery*. 1998; 228(4): 491–507

Khuri SF, Daley J, Henderson W, Barbour G, Lowry P, Irvin G, et al. 'Risk adjustment of the postoperative mortality rate for the comparative assessment of the quality of surgical care: results of the National Veterans Affairs Surgical Risk Study.' *Journal of the American College of Surgeons*. 1997;185(4): 315–327.

Khuri SF, Daley J, Henderson W, Barbour G, Lowry P, Irvin G et al. 'The National Veterans Administration Surgical Risk Study: risk adjustment for the comparative assessment of the quality of surgical care.' *Journal of the American College of Surgeons*, 1995;180(5): 519–531.

Knaus WA, Draper EA, Wagner DP, Zimmerman JE, Bergner M, Bastos PG et al. The APACHE III prognostic system. Risk prediction of hospital mortality for critically ill hospitalized adults. *Chest* 1991 (100): 1619-1636.

Knaus WA, Draper EA, Wagner DP, Zimmerman JE, APACHE II: a severity of disease classification system. *Critical Care Medicine*. 1985; (13): 818-829.

Knaus WA, Draper EA, Wagner DP, Zimmerman JE, Lawrence DE. APACHE-acute physiology and chronic health evaluation: a physiologically based classification system. *Critical Care Medicine*. 1981;9(8):591-597.

Knaus WA, Draper EA, Wagner DP, Zimmerman JE. An evaluation of outcome from intensive care in major medical centers. *Ann Intern Med*. 1986 Mar;104(3):410-8

Kohyama J, Fujitani S, Umesato Y, Kataoka H. Disease staging as a measure of disease severity. *J Medical Dental Science*. 2015;62(2):25-32.

Kozower BD, Ailawadi G, Jones DR, Pates RD, Lau CL, Kron IL et al. Predicted risk of mortality models: surgeons need to understand limitations of the University HealthSystem Consortium models. *journal of the American College of Surgeon*. 2009;209(5):551-6

Krakauer H. Office of Medical Review, Health Standards and Quality Bureau, Health Care Financing Administration. Draft Report. Baltimore, Md.: 1987. Outcomes of In-Hospital Care of Medicare Patients in 1983-1985: The Medicare Experience.

Kramer AA, Zimmerman JE: Assessing the calibration of mortality benchmarks in critical care: The Hosmer-Lemeshow test revisited. *Critical Care Medicine* 2007; 35:2052-2056.

Krumholz HM, Brass LM, Every NR, Spertus JA, Bazzarre T, Cohen DJ et al. Measuring and Improving Quality of Care. A Report from the American Heart Association/American College of Cardiology First Scientific Forum on Assessment of Healthcare Quality in Cardiovascular Disease and Stroke. American Heart Association. *Circulation*. 2000;101:1483-1493. Disponível em: <http://www.circulationaha.org>.

Krumholz HM, Lin Z, Drye EE, Desai MM, Han LF, Rapp MT, Mattera JA, Normand S-L T. An Administrative Claims Measure Suitable for Profiling Hospital Performance Based on 30-Day All-Cause Readmission Rates Among Patients with Acute Myocardial Infarction. *Circ Cardiovasc Qual Outcomes*. 2011; 4(2): 243–252.

Krumholz HM, Wang Y, Mattera JA, Wang Y, Han LF, Ingber MJ, Roman S, Normand SL. An administrative claims model suitable for profiling hospital performance based on 30-day mortality rates among patients with heart failure. *Circulation*. 2006b;113:1693–1701.

Krumholz HM, Wang Y, Mattera JA, Wang Y, Han LF, Ingber MJ, et al. An administrative claims model suitable for profiling hospital performance based on 30-day mortality rates among patients with an acute myocardial infarction. *Circulation* 2006a;113:1683-92.

Krumholz, HM. Normand, SL. Galusha, DH. Mattera, JA. Rich, AS. Wang. Risk-adjustment models for AMI and HF 30-day mortality – methodology. Washington, DC: Centers for Medicare & Medicaid Services. 2006.

Kutner NG, Ory MG, Baker DI, Schechtman KB, Hornbrook MC, Mulrow CD. Measuring the Quality of Life of the Elderly in Health Promotion Intervention Clinical Trials. *Public Health Reports*. 1992; 107(5): 530–539.

Landon B, Iezzoni LI, Ash AS, Schwartz M, Daley J, Hughes JS, et al. Judging hospitals by severity-adjusted mortality rates: the case of CABG surgery. *Inquiry*. 1996;33(2):155-66.

Lane-Fall MB. Outcomes measures and risk adjustment. *Int Anesthesiology Clinical*. 2013; 51(4): 1-10.

Laudicella A, Donni PL, Smitha PC. Hospital readmission rates: Signal of failure or success? *Journal of Health Economics*. 2013; 32: 909–921.

Lawson EH, Louie R, Zingmond DS, Brook RH, Hall BL, Han L et al. A comparison of clinical registry versus administrative claims data for reporting of 30-day surgical complications. *Annals Surgery*. 2012; 256(6):973–981

Le Gall JR, Klar J, Lemeshow S, Saulnier F, Alberti C, Artigas A, et al ICU Scoring Group.: The logistic organ dysfunction system: A new way to assess organ dysfunction in the intensive care unit. *JAMA* 1996, 276:802-810.

Le Gall JR, Lemeshow S, Saulnier F. A new Simplified Acute Physiology Score (SAPS II) based on a European/North American multicenter study. *JAMA* 1993; 270: 2957-2963.

Le Gall J-R, Loirat P, Alperovitch A, Glaser P, Granthil C, Mathieu D, et al. A simplified acute physiology score for ICU patients. *Critical Care Medicine*. 1984; 12: 975-977.

Le Gall JR. The use of severity scoring systems in the intensive care unit. *Intensive Care Medicine* 2005; 31: 1618–23.

Lemeshow S, Hosmer DW. A review of goodness of fit statistics for use in the development of logistic regression models. *Am J Epidemiology*. 1982; 115: 92-106.

Lemeshow S, Le Gall JR. Modeling the severity of illness of ICU patients. A systems update. *JAMA* 1994; 272: 1049-1055.

Lemeshow S, Teres D, Klar J, Avrunin JS, Gehlbach SH, Rapoport J. Mortality Probability Models (MPM II) based on an international cohort of intensive care unit patients. *JAMA*. 1993; 270: 2478-2486.

Lemeshow S, Teres D, Pastides H, Avrunin JS, Steingrub JS. A method for predicting survival and mortality of ICU patients using objectively derived weights. *Critical Care Medicine*. 1985; 13: 519-525.

Lindenauer PK, Bernheim SM, Grady JN, Lin Z, Wang Y, Wang Y, et al. The performance of US hospitals as reflected in riskstandardized 30-day mortality and readmission rates for Medicare beneficiaries with pneumonia. *J Hosp Med*. 2010;5(6):E12-E18.

Livingston BM, MacKirdy FN, Howie JC, Jones R, Norrie JD. Assessment of the performance of five intensive care scoring models within a large Scottish database. *Critical Care Medicine*. 2000; 28: 1820-1827.

Lohr KN, Schroeder SA. A strategy for quality assurance in Medicare. *New England Journal of Medicine* 1990; 322(10): 707–712..

Louis DZ, Gonnella JS, McCracken SB, Barnes CA. Patient classification and severity of illness: Disease Staging, DRGs, and APR-DRGs. *Proceedings of the 20th Patient Classification System/Europe Working Conference*. Budapest, Hungary. 2004; 27-30.

Louis DZ, Gonnella JS, Zeleznik C. An approach to the prevention of late hospital admissions. In *Stemming the Rising Costs of Medical Care: Answers and Antidotes*. Battle Creek, Michigan: W. K. Kellogg Foundation, 1988; 147-157.

Louis DZ, Robeson M, McAna J, Maio V, Keith SW, Liu M et al. Predicting Risk of Hospitalisation or Death: A Retrospective Population-Based Analysis. *BMJ Open*. 2014; 4: e005223. [Consultado a 10/04/2017].



Louis DZ, Yuen EJ, Braga M, Cicchetti A, Rabinowitz C, Laine C et al. Impact of a DRG-based hospital financing system on quality and outcome of care in Italy. *Health Services Research*.1999; 34 (1): 405-415.

Ma G, Hall WJ. Confidence bands for receiver operating characteristic curves. *Med Decis making* 1987; 13: 191-7.

Mainz J. Methodology matters defining and classifying clinical indicators for quality improvement. *International Journal for Quality in Health Care*. 2003;15 (6):523–530.

Manzoli L, Romano F, Schioppa FS, Staniscia T, Vitullo F, Marasco G, et al. On the use of Disease Staging for clinical management: analysis of untimely admissions in the Abruzzo Region, Italy. *Italian Journal of Public Health*. 2004; 1(3-4):103-107.

Marshall JC, Cook DJ, Christou NV, Bernard GR, Sprung CL, Sibbald WJ. Multiple organ dysfunction score: A reliable descriptor of a complex clinical outcome. *Critical Care Medicine*. 1995; 23:1638-1652.

Martin CM, Hill AD, Burns K, Chen LM. Characteristics and outcomes for critically ill patients with prolonged intensive care unit stays. *Critical Care Med* 2005; 33(9):1922-1927.

Mateus C. Portugal: Results of 25 years of experience with DRGs. *Diagnosis-Related Groups in Europe Moving towards transparency, efficiency and quality in hospitals*. European Observatory on Health Systems and Policies Series. 2011; 381 – 400.

Mateus C. O impacto do financiamento por Grupos de Diagnósticos Homogêneos em Portugal entre 1995 e 2001. Lisboa: IGIF (2002) [Consult. 04 outubro 2016]. Disponível em: [http://www.fgcasal.org/aes/docs/ceu\\_mateus](http://www.fgcasal.org/aes/docs/ceu_mateus).

Mateus C. Sistema de classificação de doentes como instrumento de gestão. In: Simões, Jorge, coord. - 30 Anos do Serviço Nacional de Saúde: um percurso comentado. Coimbra: Almedina, 2010.

Maurer PP, Ballmer PE. Hospital readmissions--are they predictable and avoidable? *Swiss Med Wkly*. 2004;134(41-42):606-11.

McCarthy EP, Iezzoni LI, Davis RB, Palmer TH, Cahalane M, Hamel MB, et al. Does Clinical Evidence Support ICD-9-CM Diagnosis Coding of Complications? *Medical Care*, 2000;38 (8), 868-876

McCrum ML, Joynt KE, Orav EJ, Gawande AA, Jha AK. Mortality for publicly reported conditions and overall hospital mortality rates. *JAMA Intern Med*. 2013; 22;173(14):1351-7.

McLoughlin V, Millar J, Mattke S, Franca M, Jonsson PM, Somekh D, Bates D. Selecting indicators for patient safety at the health system level in OECD countries. *Int J Qual Health Care*. 2006 Sep;18 Suppl 1:14-20.

McMahon LF Jr, Billi JE. Measurement of severity of illness and the Medicare prospective payment system: state of the art and future directions. *Journal of General Internal Medicine*. 1988; 3(5):482-90.

Medstat, 2010. Disease Staging Software, Version 5.27 – User Guide. Thomson Reuters, Ann Arbor, Michigan, 2010.

Mendes F. Risco: um conceito do passado que colonizou o presente: promoção da saúde. *Revista Portuguesa de Saúde Pública*. 2002; 20 (2): 53-62

Metlay JP, Schulz R, Li YH, Singer DE, Marrie TJ, Coley CM, Hough LJ, Obrosky DS, Kapoor WN, Fine MJ. Influence of age on symptoms at presentation in patients with community-acquired pneumonia. *Archives of Internal Medicine*. 1997;157(13):1453-9.

Metnitz B, Schaden E, Moreno R, Le Gall JR, Bauer P, Metnitz PG. Austrian validation and customization of the SAPS 3 Admission Score. *Intensive Care Medicine* 2009; 35: 616-622.

Metnitz PG, Lang T, Valentin A, Steltzer H, Krenn CG, Le Gall JR: Evaluation of the logistic organ dysfunction system for the assessment of organ dysfunction and mortality in critically ill patients. *Intensive Care Med* 2001, 27:992-998.

Metnitz PG, Lang T, Vesely H, Valentin A, Le Gall JR. Ratios of observed to expected mortality are affected by differences in case mix and quality of care. *Intensive Care Med* 2000; 26:1466–1472.

Metnitz PG, Valentin A, Vesely H, Alberti C, Lang T, Lenz K. Prognostic performance and customization of the SAPS II: results of a multicenter Austrian study. Simplified Acute Physiology Score. *Intensive Care Medicine*. 1999; (25): 192-197.

Metnitz PG, Moreno RP, Almeida E, Jordan B, Bauer P, Campos RA et al. SAPS 3 - From evaluation of the patient to evaluation of the intensive care unit. Part 1: Objectives, methods and cohort description. *Intensive Care Med*. 2005; 31(10):1336-1344.

Michael D. Rosko MD. DRGs and severity of illness measures: An analysis of patient classification systems. *Journal of Medical Systems*. August 1988, Volume 12, Issue 4, pp 257–274

Milic M, Goranovic T, Holjevac JK. Correlation of APACHE II and SOFA Scores with Length of Stay in Various Surgical Intensive Care Units. *Antropol*. 33 (2009) 3: 831–835

Miller ME, Hui SL, Tierney WM. Validation techniques for logistic regression models. *Stat Med*. 1991;10(8):1213-26.

Moran JL, Peisach AR, Solomon PJ, J Martin J. Cost calculation and prediction in adult intensive care: a ground-up utilization study. *Anesthesiology Intensive Care*. 2004; 32(6):787-97.

Moreno R, Vincent JL, Matos R, Mendonça A, Cantraine F, Thijs L et al. The use of maximum SOFA score to quantify organ dysfunction/failure in intensive care. Results of a prospective, multicenter study. Working Group on Sepsis related-problems of the ESICM. *Intensive Care Medicine*. 1999; 25: 686-696.

Moreno RP, Metnitz GP, Almeida E, Jordan B, Bauer P, Campos AR, et al. SAPS 3 from evaluation of the patient to evaluation of the intensive care unit. Part 2: development of a prognostic model for hospital mortality at ICU admission. *Intensive Care Med* 2005; 31:1345–1355. Moreno RP: Outcome prediction in intensive care: why we need to reinvent the wheel. *Curr Opin Crit Care* 2008, 14:483-484.

Mullin RL, Vertrees J, Freedman R, Castioni R, Tinker A. Case-mix analysis across patient populations and boundaries: A refined classification system designed specifically for international use. *3M/Health Information System*, 2002.

Myers RP, Hubbard JN, Shaheen AA, Dixon E, Kaplan GG. Hospital performance reports based on severity adjusted mortality rates in patients with cirrhosis depend on the method of risk adjustment. *Ann Hepatology*. 2012;11(4):526-35.

Naessens JM, Brennan MD, Boberg CJ, Amadio PC, Karver PJ, Podratz RO. Acquired conditions: an improvement to hospital discharge abstracts. *Qual Assur Health Care*. 1991;3(4):257-62.

Naessens JM, Huschka TR. Distinguishing hospital complications of care from pre-existing conditions. *Int J Qual Health Care*. 2004;16 Suppl 1:i27-35.

Nassar AP Jr, Mocelin AO, Nunes AL, Giannini FP, Brauer L, Andrade FM et al. Caution when using prognostic models: a prospective comparison of 3 recent prognostic models. *Journal Critical Care*. 2012; 27(4): 423.e1-7

National Center for Health Statistics (NCHS). ICD-9-CM Official Guidelines for Coding and Reporting. [Consultado a 18/07/2017]. [www.cdc.gov/nchs/data/icd9/icdguide10.pdf](http://www.cdc.gov/nchs/data/icd9/icdguide10.pdf). 2010.

National Quality Forum. Measurement Framework: Evaluating Efficiency Across Patient-Focused Episodes of care. Washington, DC: National Quality Forum, 2009. 2010b

Naved SA, Siddiqui S, Khan FH. APACHE-II score correlation with mortality and length of stay in an intensive care unit. *J Coll Physicians Surg Pak*. 2011 Jan;21(1):4-8).

Naylor MD, Brooten D, Campbell R, Jacobsen BS, Mezey MD, Pauly MV et al. Comprehensive discharge planning and home follow-up of hospitalized elders: a randomized clinical trial. *JAMA*. 1999; 281:613–620.

Naylor C, Iron K, Handa K. Measuring health system performance: problems and opportunities in the era of assessment and accountability. In: Smith, P (ed.). *Measuring up: improving health systems performance in OECD countries*. Paris: Organization for Economic Co-operation and Development. 2002.

NIH Consensus Development Conference on Critical Care Medicine. *Critical Care Med* 1983;11:466-469.

Norena M, Wong H, Thompson WD, Keenan SP, Dodek PM. Adjustment of intensive care unit outcomes for severity of illness and comorbidity scores. *Journal of Critical Care* 2006; 21(2):142 – 150.

O’Muircheartaigh C, Burke A, Murphy W. *The Index of Hospital Quality*. National Opinion Research Center at the University of Chicago (NORC) Chicago, 2004.

Palmer, R. “Process-based Measures of Quality: The Need for Detailed Clinical Data in Large Health Care Databases.” *Annals of Internal Medicine*. 1997; 127: 733-738.

Pereira J. *Economia da Saúde – Glossário de Termos e de Conceitos*. Lisboa, Associação Portuguesa de Economia da Saúde, Documento de Trabalho 1/93, Lisboa, 1993.

Peres Bota D, Melot C, Lopes Ferreira F, Nguyen Ba V, Vincent JL. The Multiple Organ Dysfunction Score (MODS) versus the Sequential Organ Failure Assessment (SOFA) score in outcome prediction. *Intensive Care Medicine*. 2002; 28: 1619-1624.

Peterson ED, Roe MT, Mulgund J, DeLong ER, Lytle BL, Brindis RG, et al. Association Between Hospital Process Performance and Outcomes Among Patients with Acute Coronary Syndromes. *JAMA*. 2006;295(16):1912-1920. [Consultado a 01/07/2017. Disponível em: <http://jamanetwork.com/pdfaccess.ashx?url=/data/journals/jama/5021/on>

Pettengill J, Vetrees J. Reliability and validity in hospital case-mix measurement. *Health Care Financing Rev*.4(2) (1982):101-128.

- Pettilä V, Pettilä M, Sarna S, Voutilainen P, Takkunen O. Comparison of multiple organ dysfunction scores in the prediction of hospital mortality in the critically ill. *Critical Care Medicine*. 2002; 30: 1705-1711
- Phelps AC, Maciejewski PK, Nilsson M, Balboni TA, Wright AA, Paulh ME et al. Religious Coping and Use of Intensive Life-Prolonging Care Near Death in Patients with Advanced Cancer. *Journal of the American Medical Association*. 2009;301(11):1140-1147
- Pine M, Fry DE, Jones B, Meimban R. Screening Algorithms to Assess the Accuracy of Present-on-Admission Coding. *Perspectives in Health Information Management* 2009; 6; 2;
- Pine M, Jordan HJ, Elixhauser A, Fry DE, Hoaglin DC, Jones B, et al. Modifying ICD-9-CM Coding of Secondary Diagnoses to Improve Risk-Adjustment of Inpatient Mortality Rates. *Med Decis Making* 2009;29:69–81.
- Pine M, Norusis M, Jones B. Predictions of hospital mortality rates: a comparison of data sources. *Ann Intern Med*. 1997;126: 347-354.
- Pinto J. — Performance organizacional: variabilidade no desempenho de dois serviços hospitalares com perfil de produção similar. Lisboa: Escola Nacional de Saúde Pública. Universidade Nova de Lisboa, 2005. Dissertação elaborada no âmbito do XXX Curso de Especialização em Administração Hospitalar 2000/ 2002 ministrado pela ENSP/UNL.
- Pisicano NJ, Veloski JJ, Brucker PC, Gonnella JS. Classifying the content of board certification examinations. *Academic Medicine*. 1989; 64: 149-154.
- Pitches DW, Mohammed MA, Lilford RJ. What is the empirical evidence that hospitals with higher-risk adjusted mortality rates provide poorer quality care? A systematic review of the literature. *BMC Health Services Research* 2007, 7:91 (<http://creativecommons.org/licenses/by/2.0>). Acedido a 25/09/2017
- Pollack MM, Ruttimann UE, Getson PR. The Pediatric Risk of Mortality (PRISM) Score. *Critical Care Medicine* 1988; 16 (11): 1110-16
- Poncet A, Perneger TV, Merlani P, Capuzzo M, Combescure C. Determinants of the calibration of SAPS II and SAPS 3 mortality scores in intensive care: a European multicenter study. *Critical Care*. 2017 Apr 4;21(1):85. doi: 10.1186/s13054-017-1673-6. Acedido a 16/10/2018.
- Pongpirula K Robinson C. Hospital manipulations in the DRG system: a systematic scoping review. *Asian Biomedicine* Vol. 7 No. 3 June 2013; 301-310.

Poole D, Rossi C, Anghileri A, Giardino M, Latronico N, Radrizzani D et al. External validation of the Simplified Acute Physiology Score (SAPS) 3 in a cohort of 28,357 patients from 147 Italian intensive care units. *Intensive Care Medicine*. 2009; (35): 1916-1924.

Portaria 163/2013, Diário da República n.º 80/2013, Série I de 2013-04-24.

Portaria 234/2015. Diário da República n.º 153/2015, Série I de 2015-08-07. A tabela preços tem por base o agrupador de GDH, All Patients Refined DRG, versão 31.0, desenvolvido nos EUA, que corresponde à versão da ICD-9-CM de outubro de 2013.

Portaria n.º 325/89 de 4 de maio. DR - I SÉRIE Nº 102, p.1820. Circular Normativa Nº 1/89.

Portugal. Ministério da Saúde. Administração Central do Serviço de Saúde. Circular normativa Nº 31/2014/DPS/ACSS de 22-12-2014: Agrupador de GDH All Patient Refined DRG.

Price L, Estrada CA, Thompson D. Administrative Data Versus Corrected Administrative Data. *American Journal of Medical Quality*. 2003; 18(1):38-45.

Quality Measurement Advisory service. Quality Measurement Tools – Assessing Hospital Performance. QMAS, Washington, 1997

Quan, H. Parsons, GA. Ghali, WA. 'Validity of information on comorbidity derived from ICD-9-CM administrative data.' *Medical Care*. 2002;40(8): 675–685.

Rao HM, Marella P, Kath B. Assessment of Severity and Outcome of critical Illness. *Indian Journal of Anaesthesia* 2008; 52(Suppl 5):652-662.

Rattner SL, Louis DZ, Rabinowitz C, Gottlieb JE, Nasca TJ, Markham FW et al. Documenting and comparing medical students' clinical experiences. *JAMA*. 2001; 256 (9): 1035-1040.

Reis Miranda D, Jegers M. Monitoring cost in the ICU: a search for a pertinent methodology. *Acta Anaesthesiol Scand*. 2012;56(9):1104-1113.

Rhodes A, Moreno RP, Azoulay E, Capuzzo M, Chiche JD, Eddleston J et al. Prospectively defined indicators to improve the safety and quality of care for critically ill patients: a report from the Task Force on Safety and Quality of the European Society of Intensive Care Medicine. *Intensive Care Med*. 2012; 38(4):598-605.

Ridley S. Severity of illness scoring system and performance appraisal. *Anaesthesia* 1998; 53: 1185 – 94.

Roffman CE, Buchanan J, Allison GT. Charlson Comorbidities Index. *Journal of Physiotherapy*. 2016;62 (3):171.

Rogers V. "Applying Inpatient Coding Skills Under Prospective Payment." AHIMA Coding Series Workbook. Chicago, IL: AHIMA, 2001.

Romano PS, Roos LL, Luft HS, Jollis JG, Doliszny K. A comparison of administrative versus clinical data: coronary artery bypass surgery as an example: Ischemic Heart Disease Patient Outcomes Research Team. *J Clin Epidemiol*. 1994;47:249 –260.

Romano PS, Roos LL, Jollis JG. Adapting a clinical comorbidity index for use with ICD-9-CM administrative data: Differing perspectives. *Journal Clinical Epidemiology* 1993;46:1075-79.

Rosen AK, Mayer-Oakes. Episodes of Care: Theoretical Frameworks Versus Current Operational Realities. *The Joint Commission Journal on Quality Improvement*. 1999; 25, (3):111-128

Rosen A, Wu J, Chang BH, Berlowitz D, Rakovski C, Ash A et al. Risk adjustment for measuring health outcomes: an application in Veterans Affairs long-term care. *American Journal of Medical Quality*. 2001;16(4):118-127.

Rosenthal GE. Weak associations between hospital mortality rates for individual diagnoses: implications for profiling hospital quality. *Am J Public Health*. 1997 Mar;87(3):429-33.

Rosenthal GE, Shah A, Way LE, Harper DL. Variations in Standardized Hospital Mortality Rates for Six Common Medical Diagnoses: Implications for Profiling Hospital Quality. *Medical Care*. 1998; 36(7): 955-964.

Rosko MD. DRGs and severity of illness measures: an analysis of patient classification systems. *Journal of Medical Systems*. 1988;12(4):257-274.

Ross JS, Normand SL, Wang Y, Ko DT, Chen J, Drye EE. Hospital volume and 30-day mortality for three common medical conditions. *N Engl J Med*. 2010;362:1110–1118

Rothen HU, Küng V, Ryser DH, Zürcher R, Regli B. Validation of "nine equivalents of nursing manpower use score" on an independent data sample. *Intensive Care Medicine*. 1999; 25: 606-611.

Rothen HU, Stricker K, Einfalt J, Bauer P, Metnitz PG, Moreno RP et al: Variability in outcome and resource use in intensive care units. *Intensive Care Med* 2007, 33:1329-1336

Rowell KS, Turrentine FE, Hutter MM, Khuri SF, Henderson WG. Use of national surgical quality improvement program data as a catalyst for quality improvement. *J Am Coll Surg* 2007;204:1293-300

Rudman WJ, Pierce W, Hart-Hester S. Healthcare Fraud and Abuse. *Perspect Health Inf Manag*. 2009; 6(Fall): 1g. Published online 2009 Sep 16.

Ruttiman UE. Statistical Approaches to Development and Validation of Predictive Instruments. *Critical Care Clinics*. 1994;10 (1), 19-35.

Santana R. O financiamento hospitalar e a definição de preços. *Revista Portuguesa de Saúde Pública* 5 (2005) 93-18.

Shahian DM, Iezzoni, Meyer GS, Kirle L, Normand ST. Hospital-wide Mortality as a Quality Metric: Conceptual and Methodological Challenges. *American Journal of Medical Quality* XX(X). 2011 by the American College of Medical Quality Reprints and permission. [Consultado a 07/01/2017]. Disponível em: <http://www.sagepub.com/journalsPermissions.nav>.

Shahian DM, Iezzoni LI, Meyer GS, Kirle L, Normand SL. Hospital-wide mortality as a quality metric: conceptual and methodological challenges. *Am J Med Qual*. 2012; 27(2):112-23

Shahian DM, Edwards FH, Ferraris VA, Haan CK, Rich JB, Normand S, et al. 'Quality measurement in adult cardiac surgery: Part 1 – conceptual framework and measure selection.' *Annals of Thoracic Surgery*, 2007;83(Suppl. 4): S3–S12.

Shahian DM, Silverstein T, Lovett AF, Wolf RE, Normand SL. 'Comparison of clinical and administrative data sources for hospital coronary artery bypass graft surgery report cards.' *Circulation*, 2007a;115(12): 1518–1527.

Shahian DM, Blackstone EH, Edwards FH, Grover FL, Grunkemeier GL, Naftel DC, et al. Cardiac surgery risk models: a position article. *Annals of Thoracic Surgery*. 2004;78(5):1868-77.

Shojania KG, Showstack J, Wachter RM. Assessing hospital quality: a review for clinicians. *Eff Clin Pract* 2001;4:82-90.

Shwartz M, Iezzoni LI, Moskowitz MA, Ash AS, Sawitz E. The importance of comorbidities in explaining differences in patient costs. *Med Care* 1996a;34:76.

Shwartz RM, Gagnon DE, Muri JH, Zhao QR, Kellogg R. Administrative Data for Quality Improvement. *Pediatrics* 1999, 103 (1): 291-301.



Shwartz M, Iezzoni LI, Ash AS, Mackiernan YD. Do Severity Measures Explain Differences in Length of Hospital Stay? The Case of Hip Fracture. *Health Serv Res.* 1996b; 31(4): 365–385.

Silber JH, Rosenbaum PR, Trudeau ME, Chen W, Zhang X, Kelz RR et al. Changes in prognosis after the first postoperative complication. *Med Care.* 2005;43(2):122-31.

Silber JH, Williams SV, Krakauer H, Schwartz JS. Hospital and patient characteristics associated with death after surgery: A study of adverse occurrence and failure to rescue. *Medical Care.* 1992; 30, 615–628.

Silva LK. Validity of the risk adjustment approach to compare outcomes. *Caderno de Saúde Pública.* 19(1) (2003): 287-295.

Sinuff T, Adhikari NK, Cook DJ, Schunemann HJ, Griffith LE, Rocker G et al. Mortality predictions in the intensive care unit: comparing physicians with scoring systems. *Critical Care Medicine.* 2006 Mar;34(3):878-85.

Siregar S, Pouw ME, Moons KGM, Versteegh MIM, Bots ML, van der Graaf Y, Kalkman CJ, van Herwerden LA, Groenwold RHH. The Dutch Hospital Standardised Mortality Ratio (HSMR) method and cardiac surgery: benchmarking in a national cohort using hospital administration data versus a clinical database. *Heart* 2014;100:702–710. Doi: 10.1136/heartjnl-2013-304645. [Consultado a 07/01/2017]. Disponível em: <http://heart.bmj.com>.

Smith P. “Registries and Care with Evidence Development” In *Clinical Data as the Basic Staple of Health Learning. Creating and Protecting a Public Good: Workshop summary*, edited by Goolsby WA, Olsen L, M McGinnis 125 – 136. Institute of Medicine Roundtable on Evidence-Based Medicine Washington, DC: National Academies Press. 2010

Stevens VG, Hibbert CL, Edbrooke DL. Evaluation of proposed casemix criteria as a basis for costing patients in the adult general intensive care unit, *Anaesthesia* 1998; 53(10):944-950.

Strand K, Flaatten H. Severity scoring in the ICU: a review. *Acta Anaesthesiol Scand.* 2008;52(4):467-78.

Thomas JE, Brennan TA. Incidence and types of preventable adverse events in elderly patients: population based review of medical records. *BMJ* Volume, 2000;320:741–4. [Consultado a 12/08/2016]. Disponível em: [www.bmj.com](http://www.bmj.com)

Thomas JW, Hofer TP. Accuracy of Risk-Adjusted Mortality Rate as a Measure of Hospital Quality of Care. *Medical Care*. 1999 37 (1), 83-92.

Thomas JW e Hofer TP. Research evidence on the validity of risk-adjusted mortality rate as a measure of hospital quality of care. *Med Care Res Rev*. 1998 Dec;55(4):371-404.

Thomas JW, Holloway JJ, Guire KE. Validating risk-adjusted mortality as an indicator for quality of care. *Inquiry*. 1993;30:6-22.

Thomas JW, Ashcraft M, Zimmerman J. An Evaluation of Alternative Severity of Illness Measures for Use by University Hospitals. Vols. 1-3. Ann Arbor, MI: Department of Health Services Management and Policy, School of Public Health, The University of Michigan, 1986.

Thombs BD, Singh VA, Halonen J, Diallo A, Milner SM. The effects of preexisting medical comorbidities on mortality and length of hospital stay in acute burn injury: evidence from a national sample of 31,338 adult patients. *Ann Surg*. 2007;245(4):629-634.

Timsit JF, Fosse JP, Troché G, De Lassence A, Alberti C, Garrouste-Orgeas M et al. Calibration and discrimination by daily Logistic Organ Dysfunction scoring comparatively with daily Sequential Organ Failure Assessment scoring for predicting hospital mortality in critically ill patients. *Critical Care Medicine*. 2002; 30:2003-2013

Tourangeau A, Tu JV. Developing risk-adjusted 30-day hospital mortality rates. *Research in Nursing & Health*. 2003; 26, 483-496. [Consultado a 13/12/2016]. Disponível em: <http://onlinelibrary.wiley>.

Tucker J. Patient volume, staffing, and workload in relation to risk-adjusted outcomes in a random stratified sample of UK neonatal intensive care units: a prospective evaluation. *Lancet* 2002;359:99–107

Urbano J, Bentes M. Definição da produção do hospital: os grupos de diagnósticos homogêneos. *Revista Portuguesa de Saúde Pública* 8:1 (1990) 49-60.

Urbano J, Bentes M, Vertrees JC. National commitment and the implementation of DRGs, in J.R. Kimberly, G. de Pouvourville et al., eds. *The Migration of Managerial Innovation*. 1993; San Francisco, CA: Jossey-Bass Inc.

Valderas JM, Starfield B, Sibbald B, Salisbury C, Roland M. Defining comorbidity: implications for understanding health and health services. *Annals of Family Medicine*. 2009;7(4):357-63.

Vincent JL, Carvalho FB. Severity of illness seminars respiratory and critical care medicine 2010; 31(1): 31-38

Valente MC. Contratualização em contexto hospitalar. *Revista nacional de Saúde Pública*. 9 (2010) 25-39.

Valentin A, Ferdinande P, ESICM Working Group on Quality Improvement. Recommendations on basic requirements for intensive care units: structural and organizational aspects. *Intensive Care Med*. 2011;37(10):1575-87

Vincent JL, Creteur J. Concept: intensive care units (ICUs) and intermediate care units. *Organization and Management of Intensive Care*. European Society of Intensive Care Medicine. Brussels, Belgium 2010.

Vincent JL, Moreno R. Clinical review: scoring systems in the critically ill. *Critical Care*. 2010; 14(2):207.

Vincent JL, Moreno R, Takala J, Willatts S, De Mendonça A, Bruining H et al. The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. On behalf of the Working Group on Sepsis-Related Problems of the European Society of Intensive Care Medicine. *Intensive Care Medicine*. 1996; 22: 707-710.

Vincent JL, Opal SM, Marshall JC. Ten reasons why we should not use severity scores as entry criteria for clinical trials or in our treatment decisions. *Critical Care Medicine*. 2010; 38:283-287.

Wagner DP, Draper EA. Acute physiology and chronic health evaluation (APACHE II) and Medicare reimbursement. *Health Care Financ Rev*. 1984; (Suppl): 91–105.

Wagner DP, Knaus WA, Draper EA. Statistical validation of a severity of illness measure. *American Journal Public Health*. 1983;73(8):878-884.

Weingart SN, Mukamal K, Davis RB, Davies Jr DT, Palmer RH, Chalane M, et al. Physician-Reviewers' Perceptions and Judgements about Quality of Care. *International Journal for Quality in Health Care*, 2001; 13 (5), 357-365.

Weingart SN, Iezzoni LI, Davis RB, Palmer RH, Cahalane M, Hamel MB, et al. Use of administrative data to find substandard care: validation of the complications screening program. *Medical Care*. 2000;38(8):796-806.

Wong DT, Knaus WA. Predicting outcome in critical care: the current status of the APACHE prognostic scoring system. *Can J Anaesth* 1991; 38: 374– 83.

Wong J, Taljaard M, Forster AJ, Escobar GJ. Van Walraven. Derivation and validation of a model to predict daily risk of death in hospital. *Medical Care*. 2011;49, (8): 734-743

Young WW. Incorporating severity of illness and comorbidity in case-mix measurement. *Health Care Financing Review*. 1984;(Suppl): 23–31.

Zhan C, Elixhauser, A, Friedman, B, Houchens, R, Chiang, YP (2007). 'Modifying DRG-PPS to include only diagnoses present on admission: financial implications and challenges.' *Medical Care*. 2007;45(4): 288–291.

Zimmerman JE, Wagner DP, Draper EA, Wright L, Alzola C, Knaus WA. Evaluation of acute physiology and chronic health evaluation III predictions of hospital mortality in an independent database. *Critical Care Medicine*. 1998; 26: 1317-1326.

Zygun DA, Laupland KB, Fick GH, Sandham JD, Doig CJ. Limited ability of SOFA and MOD scores to discriminate outcome: a prospective evaluation in 1,436 patients. *Can J Anaesth* 2005; 52: 302-308.

## 8. ANEXOS

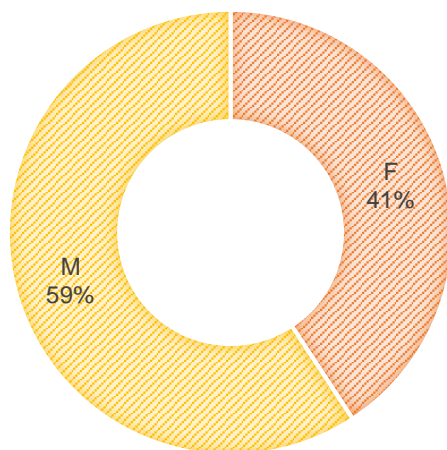
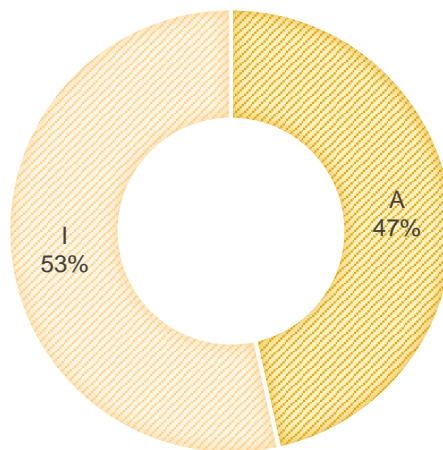


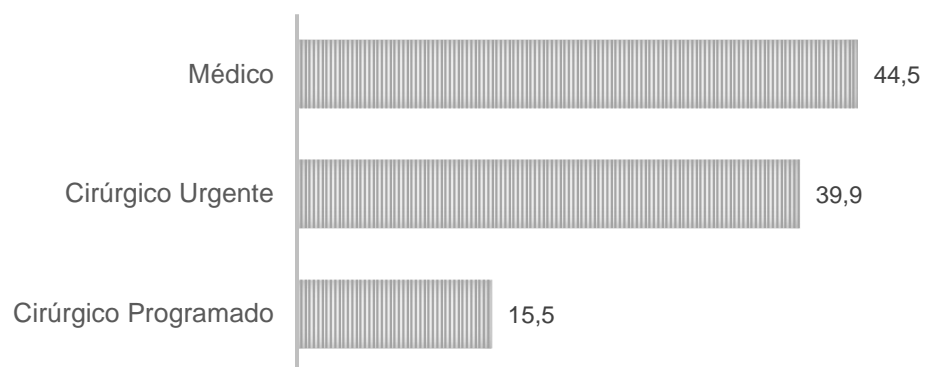
GRÁFICO 12 DISTRIBUIÇÃO DO VARIÁVEL GÊNERO



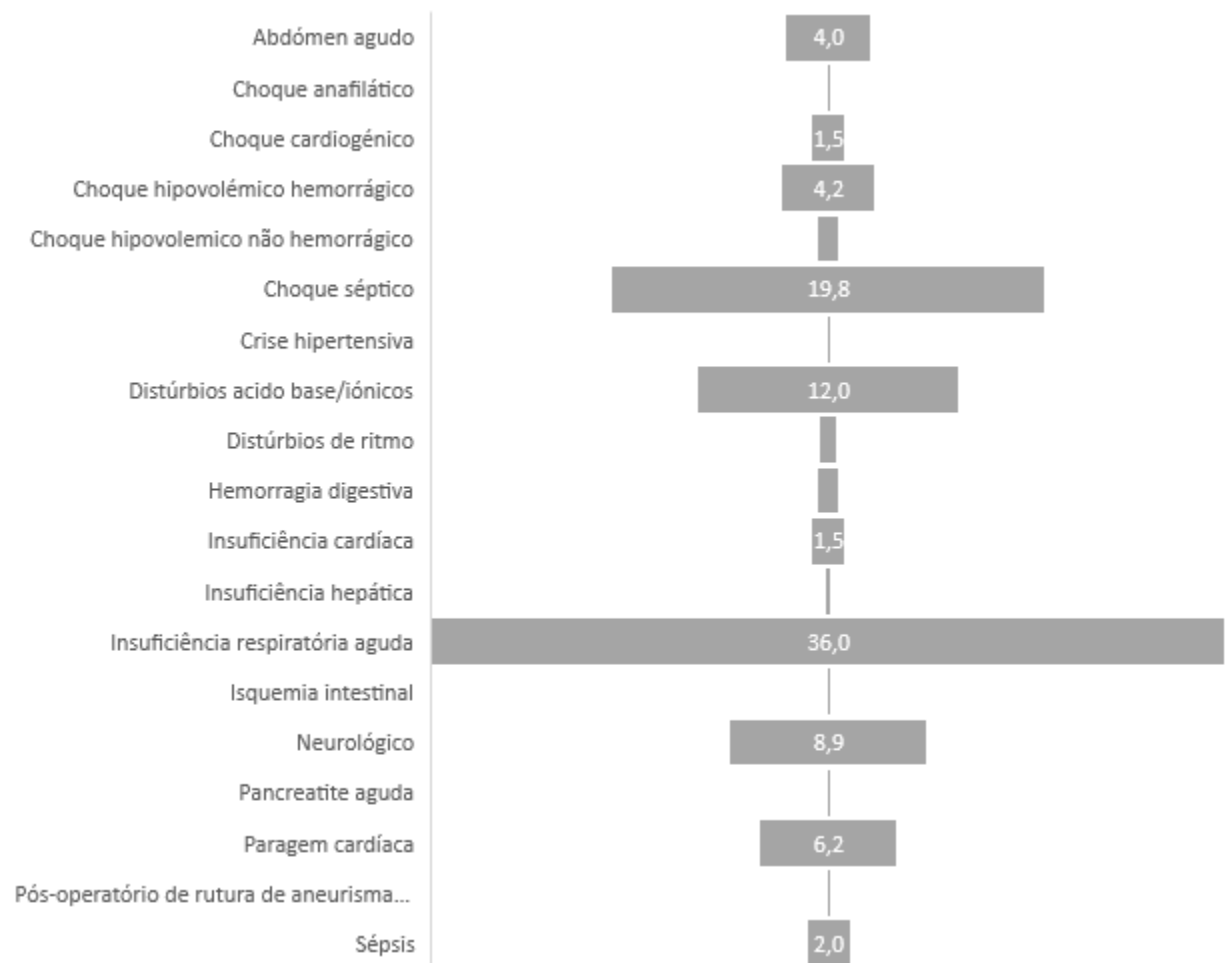
GRÁFICO 13 TIPO DE ADMISSÃO URGENTE VS PROGRAMADA



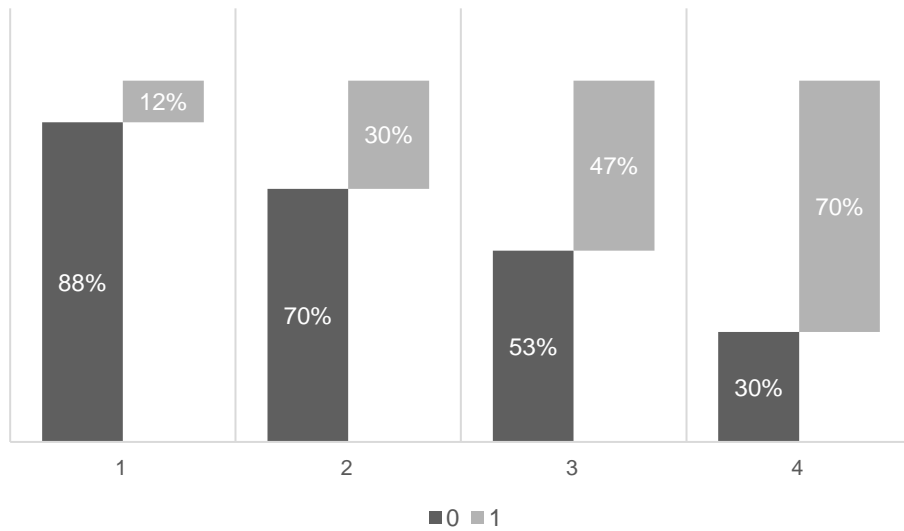
**GRÁFICO 14 ORIGEM DO DOENTE EM PERCENTAGEM URGÊNCIA VS INTERNAMENTO**



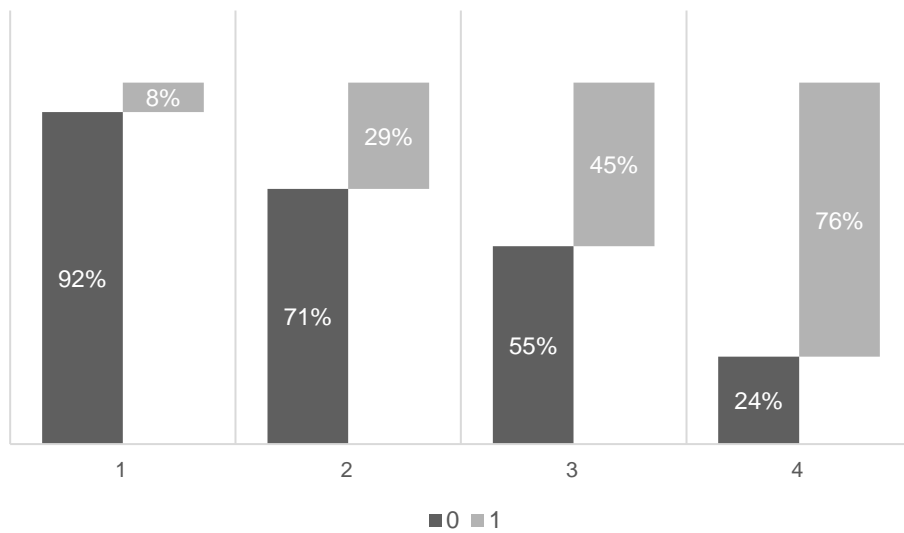
**GRÁFICO 15 PERCENTAGEM POR TIPO DE DOENTE**



**GRÁFICO 16 DIMENSÃO EM PERCENTAGEM POR MOTIVO DE ADMISSÃO**

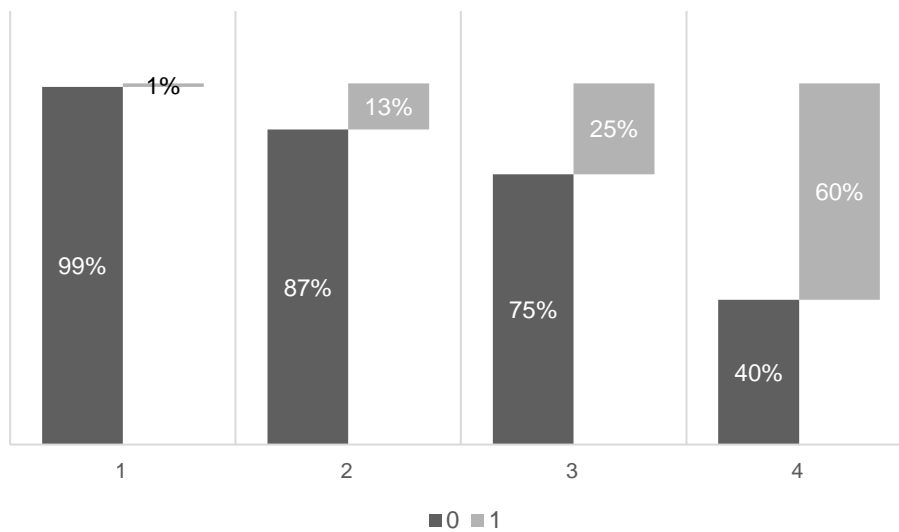


**GRÁFICO 17 PERCENTIS VS VARIÁVEL OUTCOME PARA O SISTEMA APACHE II**

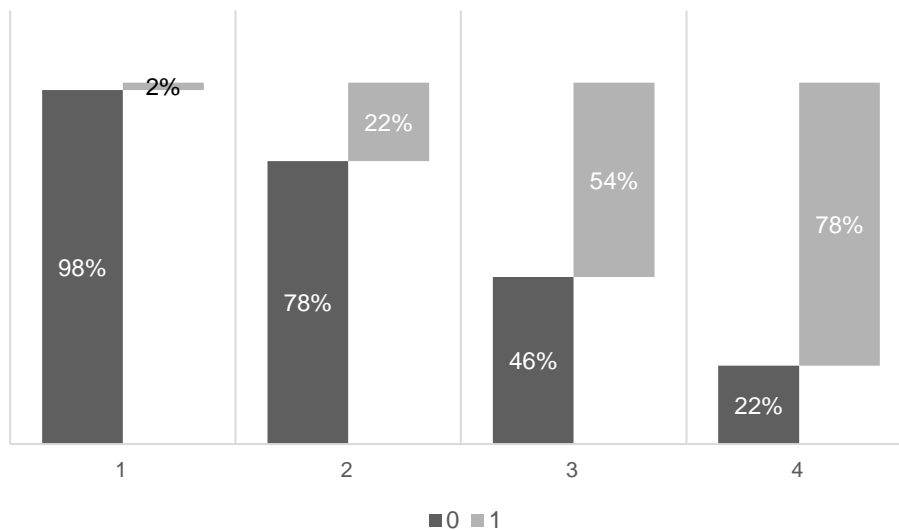


**GRÁFICO 18 PERCENTIS VS VARIÁVEL OUTCOME PARA O SISTEMA SAPS II**





**GRÁFICO 19 PERCENTIS VS VARIÁVEL OUTCOME PARA O SISTEMA APR-DRG**



**GRÁFICO 20 PERCENTIS VS VARIÁVEL OUTCOME PARA O DISEASE STAGING**

**TABELA 26 EPISÓDIOS IDENTIFICADOS COMO OBSERVAÇÕES MAL CLASSIFICADAS EM CADA SISTEMA**

(1) APACHE II	(2) APACHE II	(3) SAPS II	(4) SAPS II	(5) SAPS II	APR-DRG	APR-DRG	Dstaging
13012980	13017050	13006534	13006534	13006534	13028570	14051986	14003899
13017050	13017270	13007976	13016967	13007741	15001314	15003377	14011191
13017270	13017737	13009087	13017050	13014981	13001596	14038050	14026440
13017737	13026236	13016967	13021187	13016967	13023146	15006686	14028324
13026236	13031214	13024867	13022685	13017050	13027711	14037191	14033752
13031214	13039055	13031214	13024867	13021187	13046404	14005574	14036070
13034972	13047634	14005107	13026236	13022685	13003475	15034421	14037069
13047634	14000080	14007122	13031214	13024867	15041522		14051986
13049659	14005107	14009395	14005107	13026236	14036070		15005413
14005107	14011412	14011655	14011412	13031214	14028031		15031113
14011412	14022275	14019589	14019589	14005107	15006310		15032809
14016231	14026001	14022275	14022275	14009395	14012328		15037534
14022275	14033838	14028722	14031231	14011412	14003899		
14033838	14043165	14034016	14033838	14019589	15016895		
14036070	14047349	14043165	14034016	14022275	15029652		
14043165	14050763	14046710	14043165	14031231	15039665		
14048196	15002784	14047349	14047349	14033838	14021486		
14050763	15003377	14048007	14051762	14034016	14043165		
15002784	15003419	14051762	15001911	14043165	14046710		
15003377	15006310	15001911	15002784	14047349	13023729		
15003419	15008342	15003419	15003419	14051762	13040686		
15006310	15009198	15008342	15006310	15002784	14023712		
15008342	15015254	15036388	15008342	15003419	13048463		
15009198	15022235	15043431	15006310		14034812		
15022235	15025796	15046407	15008342		15010831		
15025796	15032031	15046427	15009198		13035461		
15032031	15034421	15009198	15031088		13027735		
15037534	15037534	15034421	15034421		12032806		
15044010	15046407	15043431	15034739		13006534		
15046407	15047593	15044010	15044010		14001827		
15047593		15046407	15046407		14018286		
		15046427	15046427		14038167		
		15047593	15047593		15019638		